PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure

Fundamental limits of high-dimensional estimation

A stroll between statistical physics, probability and random matrix theory

Soutenue par
Antoine Maillard
Le 30 août 2021

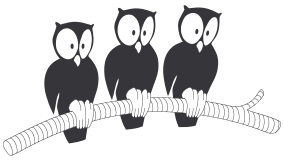
École doctorale n°564
Physique en Ile-de-France

Spécialité
Physique Théorique

Composition du jury :

Florent Krzakala ENS & EPFL	*Directeur de thèse*
Jean-Philippe Bouchaud CFM	*Rapporteur, Président du jury*
Andrea Montanari Stanford University	*Rapporteur*
Alice Guionnet ENS Lyon	*Examinatrice*
Afonso Bandeira ETH Zurich	*Examinateur*
Lenka Zdeborová EPFL	*Invitée*

ENS | PSL



The best that most of us can hope to achieve in physics is simply to misunderstand at a deeper level.

Wolfgang Ernst Pauli.

Yahaha! You found me!

Anonymous Korok seed.

Acknowledgements

For this dissertation – First of all, I would like to thank all members of the jury, especially Andrea Montanari and Jean-Philippe Bouchaud for agreeing to act as referees for this manuscript, as well as Afonso Bandeira and Alice Guionnet.

Pour la relecture, les commentaires, les suggestions, je voudrais remercier Florent, Lenka, Marc, Gérard, Alia, mais aussi ma tante Catherine pour ses suggestions concernant mon introduction, données autour d'un café.

Pour tout le reste – Tout d'abord, je voudrais remercier très chaleureusement Florent et Lenka pour ces trois années, et les innombrables discussions, rencontres et apprentissages qu'ils m'ont offerts (sans oublier Aléna et Julie qui contribuent toujours à clarifier une discussion scientifique quand elles sont présentes), et je mesure la chance que j'ai eu de travailler avec vous.

Merci à Gérard pour m'avoir fait découvrir autant de choses, et donné le goût des jolis problèmes mathématiques souvent cachés derrière ce que disent les physiciens. Merci à Marc et à Giulio, c'est toujours incroyable de discuter et de s'inspirer de deux physiciens aussi extraordinaires. Guilhem et Frédéric, ce fut un plaisir de donner avec vous le cours de mathématiques du département de physique. Merci aussi à tous les membres du groupe et au delà, anciens et plus nouveaux, pour les collaborations et les bons moments passés ensemble: Marylou, Laura, Alia, Jonathan, Benjamin, Andre, Christian, Sebastian, Maria, Francesca, Stefano, Bruno, Federica, Stéphane, Cédric, Hugo, Luca, Jean, Antoine B., et Ruben.

I am also grateful to Yue Lu for many inspiring discussions on the problem of phase retrieval while he was at ENS. My sincere thanks also go to Afonso for inviting me to ETH, and I look very much forward to working in Zurich in the coming years.

Merci également à la fondation CFM, et plus particulièrement à Nathalie, pour avoir grandement facilité les aspects plus pratiques de ces trois ans de thèse, et à Christine, de l'ENS, pour les mêmes raisons.

Pour leur amitié, leur soutiens, et les bons moments, je remercie grandement Alice, Pierre, Gaëtan (et Mailys bien sûr), Louise, Ramy, et tous les autres amis de Chambres Ski et d'ailleurs. Plus particulièrement merci aux colocs de ces trois années de m'avoir supporté: Sebastian, Nico, Nico Jack, et Nico Christelle.

A ma deuxième famille, un grand merci pour les soutiens, les accueils et toute l'affection que vous m'avez porté: merci tout d'abord à Maisaa, mais aussi à Iyad, à Teta Om Safwan, à Salam, à Anfal, à Mamdouh, à Reyhane et Mayssane pour avoir égayé de nombreuses journées, mais aussi à tous les cousins Jaber. Enfin, un énorme merci à mes deux Grand Pères, à Mamie et Bonne Maman, à Maman, Papa, Edouard, et maintenant Orianne, pour m'avoir permis d'arriver aussi loin, et m'avoir donné le goût de la science. Un grand merci aussi à tous mes frères et à mes nouvelles soeurs: Alexis, Arnaud, Augustin, Audoin, Ambroise, Lina, et Rosalie.

Enfin, je ne peux pas finir ces remerciements sans mentionner encore Alia. Je ne sais pas comment énoncer toute ma gratitude pour tout ce que tu m'as apporté jusque là (et ce que cette thèse te doit!) mais sache simplement que je ne souhaite rien de plus que te l'exprimer tous les jours qu'il me reste, et au delà.

Foreword

Throughout the thesis, the works to which I contributed are cited in orange while other references are cited in green. For instance I contributed to [Mai21] but not to [Bor14].

This thesis presents some of the work I did during my time as a Ph.D. student in École Normale Supérieure de Paris, between September 2018 and September 2021, under the supervision of Florent Krzakala (and very often also Lenka Zdeborová). It is divided into three parts, each taking a distinctive view on high-dimensional inference and learning problems.

Organization of the manuscript and summary of contributions

In Part I we introduce a large set of tools, mostly originated from the statistical physics of disordered systems, that are designed to study the fundamental properties of inference problems. Chapter 1 is quite introductory and pedagogical. While it does not present any novel results, it gives a detailed introduction to the statistical physics of inference problems as a whole, and introduces many concepts and definitions that will be used throughout the thesis. It also recalls some important historical connections between statistical physics, probability, learning and random matrix theory. Chapters 2 and 3 present the main results of two publications:

[MFC+19] *High-temperature expansions and message-passing algorithms.* **Antoine Maillard**, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mézard and Lenka Zdeborová. Journal of Statistical Mechanics: Theory and Experiment 2019 (11), 113301.

[MFK+21] *Towards exact solution of extensive-rank matrix factorization.* **Antoine Maillard**, Laura Foini, Florent Krzakala, Marc Mézard and Lenka Zdeborová. In preparation (2021).

The approach of these chapters is in essence an "old-school" strategy of statistical physics for disordered systems, namely the derivation of the TAP equations via high-temperature expansions. In Chapter 2 we show how such expansions are related to modern approximations and algorithms in inference problems, while in Chapter 3 we use them to derive systematic corrections to previous incorrect approximations that aimed to solve extensive-rank matrix factorization.

Part II is divided intro three chapters, each presenting a detailed study of the fundamental limits (both information-theoretic and algorithmic) of different high-dimensional inference and learning problems, leveraging many of the tools introduced in Part I. Beyond the heuristic techniques of statistical physics, we also provide proofs of their predictions, and detailed algorithmic studies. In particular we discuss the existence of computational-to-statistical gaps (also known as hard phases) in which a problem can be solved information-theoretically but not with any polynomial-time algorithm. In Chapter 4 we discuss learning in a model of two-layers neural network known as the committee machine. It is based on the following publication:

[AMB+19] *The committee machine: Computational to statistical gaps in learning a two-layers neural network.* Benjamin Aubin, **Antoine Maillard**, Jean Barbier, Florent Krzakala, Nicolas Macris and Lenka Zdeborová. Journal of Statistical Mechanics: Theory and Experiment 2019 (12), 124023.

In Chapter 5, we show how the knowledge of data structure can be used to enhance recovery of signals in a generic inference problem known as spiked matrix estimation. It gives a detailed account of the main results of:

[ALM+20] *The spiked matrix model with generative priors.* Benjamin Aubin, Bruno Loureiro, **Antoine Maillard**, Florent Krzakala and Lenka Zdeborová. IEEE Transactions on Information Theory, 2020.

Finally, Chapter 6 focuses on phase retrieval, a particularly relevant problem for inference and optimization, and which possesses many applications across different scientific fields. We add to our study of the fundamental limits of phase retrieval a constructive derivation of a class of optimal spectral methods. This chapter allows to put many previous results of the literature into a common framework, and its conclusions are published in the following references.

[MLKZ20] *Phase retrieval in high dimensions: Statistical and computational phase transitions.* **Antoine Maillard**, Bruno Loureiro, Florent Krzakala and Lenka Zdeborová. Advances in Neural Information Processing Systems, 33 (2020).

[MKLZ21] *Construction of optimal spectral methods in phase retrieval.* **Antoine Maillard**, Florent Krzakala, Yue M. Lu and Lenka Zdeborová. Mathematical and Scientific Machine Learning, 2021.

Part III, the last one of the thesis, takes a somehow more direct approach to the problem of learning and optimization in high dimensions. It develops a framework based on the Kac-Rice formula, an important tool of random differential geometry, to understand the topology of the high-dimensional landscapes that are optimized by learning algorithms. Chapter 7 presents a work done in collaboration with Profs. Ben Arous (NYU) and Biroli (ENS) in which we derive explicit formulas for the *complexity* of these landscapes, i.e. the statistics of their number of critical points.

[MBAB20] *Landscape complexity for the empirical risk of generalized linear models.* **Antoine Maillard**, Gérard Ben Arous and Giulio Biroli. Mathematical and Scientific Machine Learning 287-327, 2020.

We end this thesis in Chapter 8 with a more direct excursion into the realm of random matrix theory. We study the large deviations of the extreme eigenvalues of a large class of random matrices, with direct applications to the topological approach to high-dimensional landscapes described above. It is based on the letter:

[Mai21] *Large deviations of extreme eigenvalues of generalized sample covariance matrices.* **Antoine Maillard**. EPL (Europhysics Letters) 133 (2), 20005, 2021.

Topics not covered in this dissertation

The present manuscript does not cover one of my PhD publications:

[BMMK18] *The mutual information in random linear estimation beyond iid matrices.* Jean Barbier, Nicolas Macris, **Antoine Maillard** and Florent Krzakala. 2018 IEEE International Symposium on Information Theory (ISIT), 1390-1394.

In this contribution we study linear estimation problems. We go beyond the restrictive i.i.d. matrix assumption and discuss the formula proposed by [TUK06] and later by [TCVS13] who used the heuristic replica method of statistical physics (that we shall encounter several times in this thesis). Using an adaptive interpolation method and random matrix theory, we prove this formula for a relevant large sub-class of rotationally invariant matrices. The techniques used in this work share similarities with the ones of [AMB+19, MLKZ20], and we introduce most of them in Chapters 4 and 6. For this reason, their application to random linear estimation is not presented in this thesis.

Notations and abbreviations

Index of abbreviations

AMP	Approximate Message Passing
BP	Belief Propagation
CDF	Cumulative Distribution Function
EC	Expectation Consistency
EP	Expectation Propagation
ERM	Empirical Risk Minimization
ESD	Empirical Spectral Distribution
GAMP	Generalized Approximate Message Passing
GLM	Generalized Linear Model
GOE/GUE	Gaussian Orthogonal/Unitary Ensemble
LDP	Large Deviation Principle
LSD	Limiting Spectral Distribution
MCMC	Monte Carlo Markov Chain
MLE	Maximum Likelihood Estimator
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
PCA	Principal Component Analysis
PDE	Partial Differential Equation
PDF	Probability Density Function
PGY	Plefka-Georges-Yedidia
RS	Replica Symmetric
($k-$)RSB	(k-th level of) Replica Symmetry Breaking
F-RSB	Full Replica Symmetry Breaking
SE	State Evolution
SK	Sherrington-Kirkpatrick
TAP	Thouless-Anderson-Palmer
(G-)VAMP	(Generalized) Vector Approximate Message Passing

Mathematical notations

An important remark – Following standard notation in random matrix theory, we will often define a parameter $\beta = 1$ for real variables (and then denote $\mathbb{K} = \mathbb{R}$) and $\beta = 2$ for complex variables (with $\mathbb{K} = \mathbb{C}$). **This β parameter must not be confused with the inverse temperature in statistical physics notations[1], for which we will rather use the notation η.** When we consider real variables (i.e. $\beta = 1$) without any possible ambiguity, we will often remove the β and $\mathbb{K}$ indications.

$x, \mathbf{x}, \mathbf{\Phi}$	Scalar, vector, matrix.
$\mathbf{x} \cdot \mathbf{y}$ or $\mathbf{x}^\intercal \mathbf{y}$	Dot product between $\mathbf{x}$ and $\mathbf{y}$.
$\|\mathbf{x}\|$	l_2 norm of $\mathbf{x} \in \mathbb{K}^n$, $\|\mathbf{x}\|^2 = \sum_{i=1}^n \|x_i\|^2$.
$\|\mathbf{Q}\|_F$	Frobenius norm of the matrix $\mathbf{Q}$, $\|\mathbf{Q}\|_F^2 = \sum_{i,j} \|Q_{ij}\|^2$.
$\mathbb{R}_+$, $\mathbb{R}_+^\star$	Set of non-negative and strictly positive reals.
$\mathbb{C}_+$	Complex numbers with strictly positive imaginary part.
$x = \Theta(y)$	Two variables of the same order, i.e. $x = \mathcal{O}(y)$ and $y = \mathcal{O}(x)$.
I_n	The identity matrix of size n.
$\mathbf{1}_n$	The vector in $\mathbb{R}^n$ with all components equal to 1.
$\mathbb{E}$	Expectation with respect to all involved random variables.
$\mathbb{E}_{X,Y}$	Expectation with respect to X, Y only.
$\langle \cdot \rangle$	Expectation with respect to the Gibbs-Boltzmann measure.
$a \overset{\mathrm{d}}{=} b$	Shortcut for a and b having same probability distribution.
$\overset{\mathrm{d}}{\to}$, $\overset{\mathrm{p}}{\to}$, $\overset{\mathrm{a.s.}}{\to}$	Limit in the weak, probability, almost sure sense.
$\mathbb{S}_\beta^{n-1}$	Unit sphere in $\mathbb{K}^n$.
$\mathbb{S}_\beta^{n-1}(R)$	Sphere in $\mathbb{K}^n$ of radius $\|\mathbf{x}\| = R$.
$\mathcal{N}_\beta(\mu, \sigma^2)$	Gaussian distribution on $\mathbb{K}$ such that $\mathbb{E}z = \mu$ and $\mathbb{E}\|z - \mu\|^2 = \sigma^2$. If $\beta = 2$ (complex variables), we also impose $\mathbb{E}(z - \mu)^2 = 0$.
$\mathcal{M}_1^+(E)$	Set of probability measures on E.
$D_{\mathrm{KL}}(\mu\|\nu)$	Kullback-Leibler divergence (or relative entropy) of $\mu, \nu \in \mathcal{M}_1^+(E)$.
$\mathcal{H}_n(\mathbb{K})$ or $\mathcal{S}_n(\mathbb{K})$	Set of symmetric ($\mathbb{K} = \mathbb{R}$) or Hermitian ($\mathbb{K} = \mathbb{C}$) matrices of size n.
$\mathcal{H}_n^+(\mathbb{K})$ or $\mathcal{S}_n^+(\mathbb{K})$	Positive symmetric ($\mathbb{K} = \mathbb{R}$) or Hermitian ($\mathbb{K} = \mathbb{C}$) matrices of size n.
$\mathcal{U}_\beta(n)$	The (compact) group of orthogonal/unitary matrices.
$\mathcal{S}_\nu$, $\mathcal{R}_\nu$, $\cdots$	Transforms of the probability measure ν (e.g. $\mathcal{S}_\nu(x) \equiv \int \nu(\mathrm{d}t)/(t - x)$).
$\mathcal{S}_{\mathbf{\Phi}}$, $\mathcal{R}_{\mathbf{\Phi}}$, $\cdots$	Transforms of the ESD of $\mathbf{\Phi} \in \mathcal{H}_n(\mathbb{K})$.
ReLU	Rectified Linear Unit, i.e. $\mathrm{ReLU}(x) = \max(0, x)$.

[1] We will always clarify possible ambiguities between these concepts.

Contents

Introduction **1**

I Statistical physics approach to inference models and neural networks **5**

1 The statistical physics toolbox **7**
1.1 Elements of Bayesian statistical inference . . . 7
1.1.1 A motivating example: classifying data . . . 7
1.1.2 Inference problems in high dimension and the Bayes-optimal setting . . . 10
1.1.3 Generalized Linear Models . . . 12
1.1.4 Gibbs-Boltzmann measure and the free entropy . . . 12
1.1.5 Estimators . . . 13
1.2 Intuitions from the physics of spin glasses . . . 14
1.2.1 Why spin glasses? . . . 14
1.2.2 Important concepts of spin glass theory . . . 16
1.2.3 Some classical spin glass models . . . 18
1.3 Static approximations to the free energy . . . 19
1.3.1 Replica theory and replica symmetry breaking . . . 19
1.3.2 Thouless-Anderson-Palmer approach . . . 23
1.4 From physics to algorithms . . . 25
1.4.1 Belief propagation (BP) . . . 26
1.4.2 Approximate Message Passing (AMP): derivation and consequences . . . 29
1.4.3 Three approximations for non-Gaussian inference problems . . . 35
1.5 Some rudiments of probability and random matrix theory . . . 38
1.5.1 Random matrix ensembles and asymptotic spectra . . . 39
1.5.2 Large deviations . . . 43
1.5.3 High-dimensional "spherical" integrals . . . 47

2 Revisiting high-temperature expansions **51**
2.1 Organization of the chapter and main results . . . 51
2.2 Plefka-Georges-Yedidia expansion step-by-step . . . 53
2.2.1 Pedagogical derivation for a spherical SK-like model . . . 53
2.2.2 Generalization to a bipartite model . . . 58
2.3 PGY expansion for inference models . . . 59
2.3.1 PGY expansion in generic models of pairwise interactions . . . 60
2.3.2 High-temperature expansions and message-passing algorithms . . . 66
2.4 Diagrammatics and free cumulants . . . 68
2.4.1 Expectation of simple cycles and free cumulants . . . 69
2.4.2 The expectation of generic diagrams . . . 70
2.4.3 Concentration of the diagrams: a second moment analysis . . . 73
2.4.4 Higher-order moments in the diagrammatics . . . 74
2.4.5 Extension to bipartite models . . . 74
2.4.6 A note on i.i.d. matrices . . . 77

3 Towards exact solution of extensive-rank matrix factorization **79**
3.1 Introduction . . . 79
3.1.1 Definition of extensive-rank matrix factorization . . . 79
3.1.2 Organization of the chapter and summary of the results . . . 80
3.2 Critical treatment of previous approaches . . . 81
3.3 TAP equations and PGY expansion . . . 82
3.3.1 Sketch of the computation . . . 82
3.3.2 The series at order 2 and the approximation of [KKM+16] . . . 85
3.3.3 Going to higher orders: open directions . . . 85
3.3.4 Symmetric matrix factorization . . . 86

II Physics joins probability: all you need for optimal estimation
A selection of high-dimensional problems **89**

4 The physics of learning in a two-layers neural network **91**
4.1 Introduction: the committee machine . . . 91
4.1.1 Classical physics predictions . . . 92
4.1.2 Main contributions of this chapter . . . 93
4.2 Main theoretical results . . . 93
4.2.1 General probabilistic model . . . 93
4.2.2 Picking from the toolbox . . . 94
4.2.3 Main theorem: the replica-symmetric formula . . . 94
4.3 Investigating computational-to-statistical gaps . . . 97
4.3.1 Approximate Message-Passing . . . 97
4.3.2 From two to more hidden neurons, and the specialization transition . . . 98
4.4 Proof of the replica formula: adaptive interpolation . . . 101
4.4.1 Interpolating estimation problem . . . 102
4.4.2 Overlap concentration and fundamental sum rule . . . 103
4.4.3 A technical lemma and an assumption . . . 105
4.4.4 Matching bounds: adapting the interpolation path . . . 106

5 Generative models, or how to exploit the structure in the data **109**
5.1 Generative models for spiked matrix estimation . . . 109
5.1.1 Introduction: exploiting data structure . . . 109
5.1.2 Inference model: spiked matrix estimation . . . 110
5.1.3 Generative models for the data . . . 111
5.1.4 Summary of main results . . . 112
5.2 Analysis of optimal estimation . . . 113
5.2.1 Mutual information: the replica method rigorous once again . . . 113
5.2.2 Optimal performance and statistical thresholds: phase diagrams . . . 115
5.2.3 Algorithmic optimal estimation . . . 118
5.2.4 State evolution equations . . . 119
5.3 LAMP: a spectral algorithm for generative priors . . . 121
5.3.1 Linearizing the AMP equations . . . 121
5.3.2 Random matrix perspective on the spectral methods . . . 122
5.3.3 Application to real data recovery . . . 124
5.4 Random matrix analysis of the transition . . . 124
5.4.1 The bulk of eigenvalues: proof of Theorem 5.2 . . . 124
5.4.2 BBP-like transition: proof of Theorem 5.3 . . . 127

6 Phase retrieval: theoretical transitions and efficient algorithms 133
6.1 The phase retrieval problem . . . 133
6.2 Optimal estimation in GLMs with structured data . . . 136
6.2.1 Replica free entropy and how to prove it . . . 136
6.2.2 Algorithmic point of view: the G-VAMP algorithm . . . 140
6.3 Weak and perfect recovery transitions . . . 141
6.3.1 Weak recovery: beating a random guess . . . 141
6.3.2 Perfect recovery for Gaussian signals in noiseless phase retrieval . . . 142
6.3.3 Surprising consequences and open questions . . . 143
6.4 Efficient algorithms: constructing optimal spectral methods . . . 143
6.4.1 Universality of the optimal method . . . 143
6.4.2 Linearized vector approximate message passing . . . 145
6.4.3 The Bethe Hessian: TAP revisited . . . 146
6.4.4 Unifying the approaches . . . 148
6.5 Statistical and algorithmic analysis: numerical experiments . . . 149
6.5.1 Optimal algorithms and computational gaps . . . 149
6.5.2 Spectral methods: cheap and efficient . . . 151
6.5.3 Real image reconstruction . . . 154

III Towards a topological approach to high-dimensional optimization 159

7 The complexity of high-dimensional landscapes 161
7.1 Counting complexity: the Kac-Rice formula . . . 161
7.1.1 How to "count" the complexity of a landscape? . . . 161
7.1.2 The area formula . . . 163
7.1.3 The Kac-Rice formula . . . 164
7.1.4 The complexity of the pure spherical p-spin model . . . 165
7.2 Kac-Rice for inference models: main results . . . 169
7.3 Proof of the annealed complexity . . . 175
7.3.1 Applying the Kac-Rice formula . . . 175
7.3.2 The complexity at finite n . . . 175
7.3.3 Concentration and large deviations . . . 177
7.4 Towards a numerical solution? . . . 178
7.4.1 The logarithmic potential of $\mu_{\alpha,\phi}[\nu]$ . . . 178
7.4.2 Heuristic derivation of simplified fixed point equations . . . 178
7.5 The quenched complexity and the replica method . . . 180
7.5.1 Computing the p-th moment . . . 180
7.5.2 Decoupling replicas and the $p \downarrow 0$ limit . . . 183

8 An excursion to large deviations in random matrix theory 187
8.1 Why the large deviations of the eigenvalues? . . . 187
8.1.1 The landscape of generalized linear models and variants . . . 187
8.1.2 PCA for correlated data . . . 188
8.1.3 Organization of the chapter . . . 188
8.2 Large deviations of extreme eigenvalues of generalized sample covariance matrices 188
8.2.1 Some formal definitions and assumptions . . . 188
8.2.2 Main result . . . 191
8.3 Monte-Carlo simulations . . . 192
8.4 Derivation of the rate function . . . 194
8.4.1 General idea behind the method . . . 194

8.4.2 Tilting the measure: a first attempt . . . 195
8.4.3 Beyond the transition: a second tilting . . . 198
8.4.4 Going further: the complex case and the left tail of the large deviations . 199

Afterword **203**

Bibliography **205**
PhD publications . . . 205
Numerical codes of PhD publications . . . 206
Other references . . . 206

Appendices **229**

A Technicalities of the Plefka-Georges-Yedidia expansion **229**
A.1 Order 4 of the expansion for a spherical model . . . 229
A.2 Generalizations of the diagrammatics . . . 231
A.3 PGY for extensive-rank matrix factorization . . . 233
A.4 The expansion for symmetric extensive-rank matrix factorization . . . 238

B Details of replica computations **241**
B.1 Replica calculation for the committee machine . . . 241
B.2 Replica computation for generic GLMs . . . 243

C Proving the replica formula: details in the committee machine **251**
C.1 Positivity of some matrices . . . 251
C.2 Properties of the auxiliary channels . . . 251
C.3 Setting in the Hamiltonian language . . . 252
C.4 Free entropy variation: Proof of Proposition 4.3 . . . 253
C.5 A few technical lemmas . . . 255

D Technical results of Part II **257**
D.1 Generalization error in the committee machine . . . 257
D.2 Large K limit in the committee machine . . . 258
D.3 RMT analysis of the spiked matrix model . . . 263
D.4 State evolution of spectral methods with generative prior . . . 267
D.5 Derivation of thresholds in phase retrieval . . . 268
D.6 Details of the spectral methods analysis for phase retrieval . . . 272

E Details of the topological approach **275**
E.1 The quenched complexity calculation . . . 275
E.2 Details of proof for the annealed complexity . . . 278
E.3 The large deviations in the white Wishart case . . . 290
E.4 The phase transition in the rate function . . . 290
E.5 Technicalities on spherical integrals . . . 291
E.6 Simplifying the rate function . . . 295

Introduction

An unexpected journey

This thesis is an incursion between the theory of computation and theoretical physics. Computation theory aims at understanding the capabilities and limitations of computers and algorithms, not through the analysis of the performance of actual machines, but in the fundamental sense: it studies the mathematical structure of problems, in order to discern which problems are solvable and which are not, and to design algorithms capable of solving them in the most efficient possible way. The ever-growing impact of computation on our modern world has led computation theory, and theoretical computer science as a whole, to become a field of scientific research that can not be overlooked. Historically, the first well-known example of an actual algorithmic procedure is perhaps the one described by Euclid in Book VII of his *Elements*, which is still taught in elementary mathematics as the canonical way to find the greatest common divisor of two integers. The performance of Euclid's algorithm is elementary to compute for relatively small numbers, see Fig. 1, but understanding it for random or very large numbers is much more demanding. Understanding the fundamental limits of computational problems (e.g. "is it possible to find the greatest common divisor of two integers ?" – the answer is obviously positive here with Euclid's algorithm) and of the algorithms themselves (e.g. "what is the best algorithm to do so, and what is its performance?"[1]) are the two questions that guide this thesis, in the context of the inference problems described below.

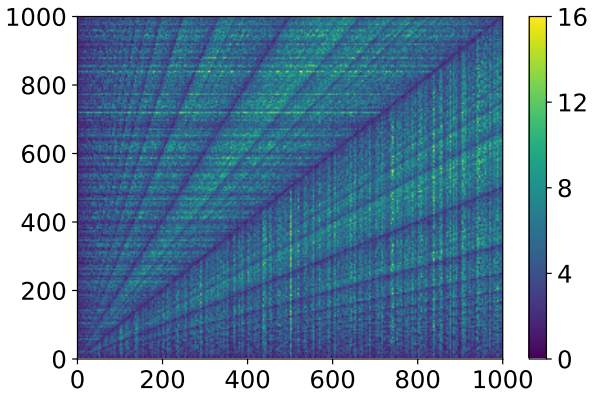


FIGURE 1: Number of operations needed by Euclid's algorithm to compute the greatest common divisor of two integers.

Physics, on the other hand, seeks to grasp some bribes of the laws of nature, from the cosmological evolution to the structure of subatomic particles. In particular, statistical physics, the branch of physics to which this thesis belongs (and perhaps one of the least-known to the general audience), deals with the fundamental properties of systems composed of a huge number of simple individual elements, whose interaction can give rise to fascinating collective behavior. Initially motivated by the study of gases, the field was developed in the 19^{th} century by the combined genius of Boltzmann, Maxwell, and Gibbs [Max60, Bol98, Gib02]. For a non-specialist audience, a statistical physicist studying computation theory might look somehow out-of-place: what could the physics of gases have to do with the fundamental limits of computational problems? This brief introduction aims at providing an intuitive description, accessible to a general audience, of this surprising association.

Entropy – A first intuition on the existence of this connection can be grasped by considering the notion of *entropy*. It is indeed foundational in both fields, albeit being formalized by Boltzmann in the 1870s in statistical mechanics and by Shannon in 1948 in theoretical computer science

[1] The answer to this question is much harder, but the original algorithm of Euclid is not optimal in terms of computational time.

[Sha48]. From the statistical mechanics point of view, since the numbers of elements is too large to follow the individual evolution of each one (e.g. molecules in the example of a gas), we rather describe the probability of a global configuration of the system. Denoting $\mathcal{C}$ the ensemble of all possible configurations, each configuration i is assigned with a probability p_i of appearance. The *entropy* of the system is then defined as:

$$S = -\sum_{i \in \mathcal{C}} p_i \ln p_i. \tag{1}$$

As shown by Boltzmann, the entropy represents the amount of *uncertainty* present in the system. To picture it, let's imagine a single molecule which has two possible configurations A and B. If the molecule is certain to be in one of the two configurations (let's say A) then $p_A = 1$ and $p_B = 0$, which leads to $S = 0$. On the other hand, if both configurations are equiprobable, $S = -(2/2)\ln 1/2 = \ln 2 > 0$: the entropy is higher because the system is less "ordered"!

In the first half of the 20$^{\text{th}}$ century, Shannon realized that eq. (1) was also suited to quantify the notion of *information* in computational problems. Taking the example of data communication, he wanted to understand if a receiver could exactly identify a message X from an observation Y that was possibly corrupted, or truncated. If possible, this allows for instance to send compressed messages much smaller than the originals without loosing any information, or to "clean" messages that were corrupted by noise. Because of the noise or compression, the observation Y obtained by the receiver can take many values $\{Y_i\}$, each with probability p_i. The entropy of eq. (1), now applied to these probabilities, is known as the *Shannon entropy*[2]. Leveraging this definition, Shannon showed a foundational theorem: he proved that the entropy is related to the absolute mathematical limit of how much one can compress a message without loosing any information. Starting from a physical concept, we somehow described the fundamental limits of a computational problem! Based on this fundamental result, Shannon's work laid the foundations of the field of *information theory*.

Statistical physics of inference and learning – The connection between statistical physics and information theory, which inspired much of this thesis, did not end at the definition of entropy. It also inherits from diverse perspectives, whose main lines can be understood from the point of view of *Bayesian inference*. In this general framework, the observer wishes to estimate a set of parameters $\mathbf{x} = (x_1, \cdots, x_n)$ from the observations of some data $\mathbf{Y} = (Y_1, \cdots, Y_m)$ (which might contain noise), while having some prior knowledge on the parameters (for instance we may know that they must correspond to a word). The Bayesian statistician then asks the question: "Given the data $\mathbf{Y}$ that I observed, what is the probability that it was generated from some given parameters $(x_1, \cdots, x_n)$"? A key realization is that, when stated this way, many problems of Bayesian inference can be seen as statistical physics models! In this transposition, the parameters $(x_1, \cdots, x_n)$ play the role of particles, which are interacting with each other. These interactions of the physical model are shaped by the knowledge acquired through the observation of the data $\mathbf{Y}$ in the original inference task. Moreover, like in statistical mechanics, the important quantities are *macroscopic*: when the number of parameters is very large, we care about the fraction of the message that we can recover, rather than the recovery of every single parameter x_i. When stating the problem in this way, we can harness much of the power of statistical physics to tackle Bayesian inference.

Looking closer, Bayesian inference problems have a somehow unusual feature compared to classical statistical physics models. As we mentioned, the observations $\mathbf{Y}$ play the role of interactions between the particles $\mathbf{x}$; however in practical inference procedures these observations are usually

[2] Shannon named it entropy after receiving a famous advice from von Neumann: "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage."

affected by some random noise: translated into our statistical physics model, this would imply a stochastic interaction between particles! While this does not make much sense in terms of gases, such models of random interactions between particles are called *spin glasses* in physics, and they arise in a variety of situations (we will discuss further the origin of these models and their physical description in Section 1.2).

A canonical example of this representation of a statistical estimation procedure as a spin glass model was described by Gardner and Derrida in the late 1980s [GD89], for an elementary algorithm known as the *perceptron*. After physicists realized that the theory of spin glasses – that was rapidly developing since the 1970s – could be leveraged to tackle many open problems in information theory, the way was open for decades of fruitful collaborations between communities of theoretical physicists, computer scientists, and probabilists.

For instance, spin glasses provide an intuitive understanding of the possible hardness of inference problems. Indeed, in many cases solving an inference task translates into minimizing the *energy* of the corresponding spin glass model. However, because of the random interactions between particles the energy landscape of a spin glass is generally very rugged, see Fig. 2. Unfortunately, practical algorithms used to minimize the energy are *local*: they generally start at a random point, and then try to guess the position of the minimum, while they can only see the portion of the landscape infinitely close to their current position. A natural strategy would be to go down for as long as possible: but given the form of the landscape this surely means that the algorithm will end up in a small well, very far from the actual minimum that solves the estimation problem. Worse, the algorithm will then have no way of knowing if it found the lowest point or not! With this picture in mind, one can get an idea of why solving the original estimation problem might be computationally hard. This is just a tiny glimpse of the incredible amount of intuition offered on inference problems by the physics of spin glasses, these "unicorns" that the prolific physicist Sam Edwards was chasing[3].

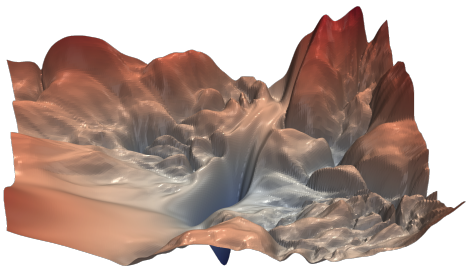


FIGURE 2: Example of a rugged landscape in dimension two. Picture taken from [LXT+18].

From the physics of spin glasses and their connections to theoretical computer science, we are now reaching the scientific questions that drive this thesis. Namely, this dissertation will discuss the following general question through the prism of the statistical physics of spin glasses:

How do statistical estimation problems behave in very high dimension ?

Here, "very high dimension" means that there are many parameters to infer, but we also have access to comparably many data. As we will detail in Section 1.1 this is the relevant hypothesis to analyze recent artificial intelligence algorithms, such as the celebrated *deep learning* techniques. These modern methods rely on the optimization of a very large

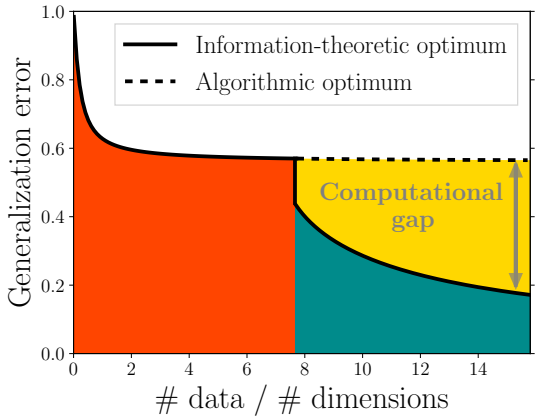


FIGURE 3: A computational gap (or hard phase) in learning in a two-layers neural network. The full line is the theoretical optimal error, while the dashed line is the error achieved by the best-known algorithm. When the number of data divided by the number of parameters exceeds a threshold (around 7.5) all algorithms are incapable of reaching the information-theoretic optimum in reasonable (i.e. polynomial) time (see Chapter 4).

[3] in the words of Pierre-Gilles de Gennes [GEG+05].

number of internal parameters (organized in what is called a *neural network* in the case of deep learning) using the gigantic amounts of data provided by our evermore inter-connected world, and are currently achieving state-of-the-art (and often super-human) performance for tasks as diverse as image classification, natural language processing, or speech recognition. In this thesis, we will be harnessing theoretical tools developed in spin glass theory to understand the fundamental limitations of a variety of learning and inference procedures in this high-dimensional limit, and we show an example of such results in Fig. 3. In a time in which the theory of learning and inference is an increasingly active and debated area of research, the variety of approaches developed in the spin glass and statistics literature allows to explore very diverse directions, and the three parts of this dissertation reflects some of these distinct perspectives.

Some further reading – The book of C. Moore and S. Mertens [MM11] proposes an amazing scientific journey through the theory of computation in general, and is a must-read for the audience interested in this field. The curious reader can also refer to the introduction of [Abb20] (for a short introduction accessible to a general public) or the general reviews [ZK16, Gab20] (for more technical and completes descriptions), which are all great presentations of the foundations of the field.

Part I

Statistical physics approach to inference models and neural networks

Chapter 1

The statistical physics toolbox

"[...] briefly, and in its most concrete form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data."

R. A. Fisher, On the Mathematical Foundations of Theoretical Statistics (1922).

Disclaimer – This first chapter defines important notions of the theory of statistical estimation. Most importantly, we introduce several theoretical tools and concepts that mainly originated in the statistical physics literature, starting from the 1970s with the seminal works on disordered systems of Edwards & Anderson [EA75] and Sherrington & Kirkpatrick [SK75]. It will provide us with a diversified toolbox to tackle the fundamental properties of inference problems in high dimension. The last Section 1.5 is more mathematical in nature, and presents several results on random matrices and large deviations theory that will prove useful throughout this dissertation.

Bibliographical note – Let us start by mentioning a few important references on the statistical physics approach to disordered systems and high-dimensional inference, for the curious reader who might finish this introductory chapter with more questions than answers. [MPV87] is an ageless work that focuses on models of disordered systems known as *spin glasses* and introduces several methods used in this thesis, e.g. the replica theory or the Thouless-Anderson-Palmer equations. Several important works followed and focused on the links between the physics of spin glasses and estimation models. Let us mention a few particularly impacting ones: [Nis01], which presents a detailed derivation of the methods, [MM09] with a point of view perhaps more adapted to theoretical computer scientists and information theorists, or [ZK16] with a specific focus on the physics approach to Bayesian inference. Finally, [Gab20] is a recent review of the theoretical understanding of learning in the context of neural networks, using a variety of "mean-field" techniques that originated in the statistical physics literature.

1.1 Elements of Bayesian statistical inference

1.1.1 A motivating example: classifying data

Before diving into more technical details, we first motivate the theoretical study of inference by the important example of *binary classification* of data.

Imagine a medical student in neuroradiology, which has seen during her long studies thousands of brain MRI images, belonging both to patients with a brain tumor and to healthy ones. One

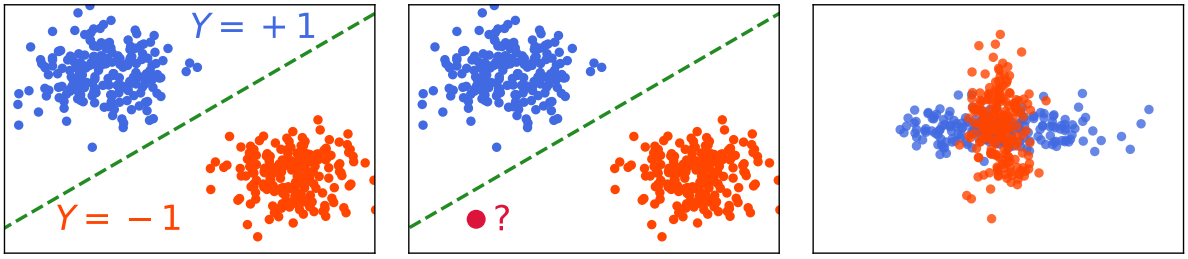


FIGURE 1.1: Classifying data points in $\mathbb{R}^2$ among two sets, represented as red and blue points. (Left) Linearly separable data, we draw a possible decision boundary learned by the perceptron. (Middle) We add a yet unseen point, for which the algorithm now predicts the label $Y = -1$. (Right) Non-linearly separable data, impossible to fit perfectly with the perceptron algorithm.

day she is for the first time confronted to the brain image of an undiagnosed patient, and must deduce by herself alone[1] if he is affected by a tumor. How can she minimize the risk of a mistake? This problem enters the category of "binary classification", and arises in several areas of statistics and computer science. A natural and systematic idea would be to classify all the images she has seen before into two categories – cancerous and non-cancerous – and to compare the new image to all these known diagnostics. After having looked through all of them, she finds the past image which is the most similar to the new one, and then takes a decision based on the category to which this known image belongs to. While very naive, this decision process is a simple instance of the *k-nearest neighbors* algorithm [CH67], which solves the binary classification problem by assigning a new data point to the label of its closest known examples.

Unfortunately the nearest-neighbor approach is quite limited, especially as it is very sensitive to noisy data or to outliers: if a few cancerous pictures are very atypical they could greatly compromise the output of the procedure! To circumvent some of these limitations we describe in the following two strategies for binary classification based on models of *neural networks*.

The perceptron – One of the historically most important techniques used to perform binary classification of data is the *perceptron* algorithm, introduced by Rosenblatt in 1957 [Ros57]. Given a set of *weights* $\mathbf{W} \in \mathbb{R}^n$ and a bias $b \in \mathbb{R}$, it predicts the classification $Y \in \{-1, +1\}$ of a data point $\mathbf{z} \in \mathbb{R}^n$ as follows:

$$Y = \text{sign}\Big(\sum_{i=1}^{n} W_i z_i + b\Big). \tag{1.1}$$

In mathematical terms, the perceptron cuts the space $\mathbb{R}^n$ in two regions at the *decision boundary*, i.e. the hyperplane characterized by the equation $\sum_{i=1}^n W_i z_i + b = 0$. The perceptron algorithm then consists in using labeled points (i.e. a set of m values $\{\mathbf{z}_\mu, Y_\mu\}$) to *train* the algorithm, that is to learn $\mathbf{W}, b$ such that eq. (1.1) matches "as best as possible" the vectors $\mathbf{z}_\mu$ to the labels Y_μ. The hope is that it will then allow to predict the class of a new point previously unseen in the training procedure, as the algorithm will have "learned" how to differentiate the two types of data points. Importantly, the perceptron algorithm can only perfectly fit a training set $\{\mathbf{z}_\mu, Y_\mu\}$ which is *linearly separable*, i.e. that can be separated in two by a decision hyperplane. We summarize this discussion in Fig. 1.1. Note that the perceptron algorithm can also be adapted to solve *regression* tasks, in which the labels are not discrete like in the present example, but continuous: one simply replaces the sign function in eq. (1.1) by a smooth function φ.

Multi-layer neural networks – In order to classify data of increasing complexity, such as natural images (which will surely not be linearly separable), a variety of subsequent algorithms

[1] (a hopefully unrealistic situation)

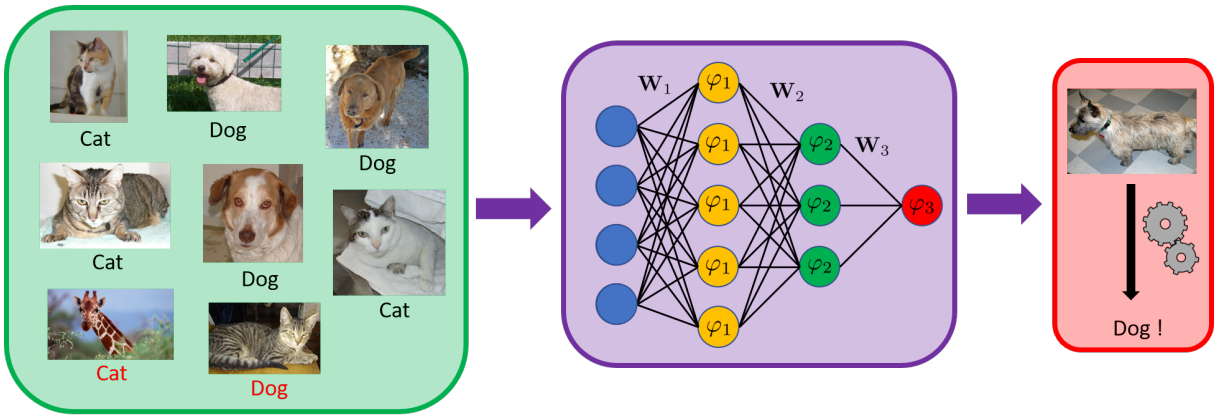


FIGURE 1.2: Classification of cats and dogs images by a fully-connected neural network. We take the training images randomly in a large set of labeled pictures (left). Note that it contains a very small number of misclassified examples (labeled in red), which we will discuss in Section 1.1.2. We use it to train a fully-connected neural network of the form of eq. (1.2) (middle). We then use eq. (1.2) with the trained weights to predict the label of an unseen image (right).

were developed. Fully-connected neural networks, also called *multi-layer perceptrons*, are among them. In these models, the input data $\mathbf{z}$ is propagated among a number $L \geq 1$ of layers, adding a non-linearity φ_l component-wise in each layer, which results in the observation:

$$Y = \varphi_L(\mathbf{W}_L \varphi_{L-1}(\cdots \varphi_1(\mathbf{W}_1 \mathbf{z}))). \tag{1.2}$$

For classifying data in two categories we simply have $\varphi_L : \mathbb{R} \to \{-1, 1\}$. By use of *training algorithms* (discussed at a later point of this chapter) on samples of data $\{\mathbf{z}_\mu, Y_\mu\}$, we infer the value of the weight vectors $\{\mathbf{W}_l\}$, so that eq. (1.2) is able to accurately predict the category of an unseen data point. The procedure is illustrated in Fig. 1.2, and we will study a neural network of this category in Chapter 4. While the introduction of the multi-layer perceptron dates back to the late 60s [MP69], it took many more years for a practical training algorithm, called backpropagation, to be developed [RHW86]. Even then, the available computing power was not sufficient to train these networks, and it was especially during the past decade that the dramatic increase of available computing power allowed to finally train networks with many layers and nodes, a strategy known as *deep learning*. Actual modern neural networks are actually much more refined than multi-layer perceptrons, and the growing range of their applications has led to a whole zoology of deep neural nets, such as convolutional neural networks (CNNs) which are particularly used in computer vision, or recurrent neural networks (RNNs) with many applications in natural language processing [LBH15]. A general theory of the learning procedure in deep neural networks however still appears beyond reach [Gab20, Zde20].

Fundamental limits of inference and learning – These sophisticated neural networks are just a particular class of *inference* procedures, which consist in extracting information (e.g. the values of the weights $\{\mathbf{W}_l\}$ in a neural network, cf. eq. (1.2)) from the analysis of data samples. Such inference procedures appear all over the scientific spectrum, as they can be applied to artificial intelligence methods, to medicine (e.g. recognizing tumors by medical imaging analysis as we discussed), self-driving systems or quantitative finance, among many. In this thesis, we consider the *fundamental limits* of inference procedures. Our main interest will therefore be to study the following question, which essentially sums up the task given to our medical student:

How much information is it possible to extract from a given set of data ?

This question of optimal performance is in a sense two-dimensional: one can consider both a *statistical* version ("How much information can we extract in principle ?") and an *algorithmic*

one ("How to extract this information? Can it be done efficiently?"), and this dissertation is concerned with both aspects. Importantly, we will consider these questions in the framework of the current "big data" age, in which computing systems have access to effectively unlimited data material, and use it to estimate a gigantic number of internal parameters.

1.1.2 Inference problems in high dimension and the Bayes-optimal setting

Worst-case and typical inference

A large part of statistical learning theory has focused on obtaining uniform bounds for the performance achieved by various estimation methods, using tools such as the Rademacher complexity or the Vapnik–Chervonenkis dimension [Vap13, AAKZ20]. Although very powerful and generic, such uniform bounds can be vacuous in the presence of very rare events, or structured data distributions. Indeed one would ideally want to understand the optimal achievable performances for *typical* realizations of the data, which are more likely to accurately describe real-world applications. The notion of typicality is natural, as one generally has access to data generated from an unknown underlying probability distribution: atypical events would then correspond to very rare realizations of the data, which we can (quite naively) illustrate by the red labels in Fig. 1.2. This has motivated another approach, that aimed at understanding the performance of said methods in the typical case, and which will be a key setting of this dissertation.

Bayes' theorem

To study typical inference, we will take the point of view of *Bayesian estimation*, which originates from the works of Bayes and Laplace in the 18$^{\text{th}}$ century [Bay63, Lap74]. This approach to inference is centered around Bayes' rule: when trying to infer a vector $\mathbf{x}$ from a set of data $\mathbf{Y}$ we can write

$$\mathbb{P}(\mathbf{x}|\mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y}|\mathbf{x})}{\mathbb{P}(\mathbf{Y})}\mathbb{P}(\mathbf{x}). \tag{1.3}$$

The left-hand side of this equation is called the *posterior distribution* of $\mathbf{x}$ given $\mathbf{Y}$. The probability $\mathbb{P}(\mathbf{x})$ is known as the *prior*: it quantifies the knowledge we have of $\mathbf{x}$ before receiving any data. Finally, $\mathbb{P}(\mathbf{Y}|\mathbf{x})$ is the *channel* distribution: it describes how the data is generated given the parameters $\mathbf{x}$.

Denoising a vector – To illustrate these concepts, imagine that we observe a real vector $\mathbf{Y} = \mathbf{X}^* + \sqrt{\Delta}\mathbf{Z}$ with $\mathbf{Z} \sim \mathcal{N}(0, \mathrm{I}_n)$: we wish to "denoise" the vector and recover $\mathbf{X}^*$. Our prior knowledge on $\mathbf{X}^*$ (independently of the observation of any data) is modeled by the prior distribution $\mathbb{P}(\mathbf{X}^*)$. Given a set of observations $\mathbf{Y}$ generated by an unknown vector $\mathbf{X}^*$, the probability that it was actually generated by $\mathbf{x}$ is:

$$\mathbb{P}(\mathbf{x}|\mathbf{Y}) = \frac{1}{\mathbb{P}(\mathbf{Y})}\frac{\exp\left\{-\frac{1}{2\Delta}\|\mathbf{Y}-\mathbf{x}\|^2\right\}}{(2\pi\Delta)^{n/2}}\mathbb{P}(\mathbf{x}).$$

Bayesian inference consists in leveraging $\mathbb{P}(\mathbf{x}|\mathbf{Y})$ to recover information about the unknown "ground-truth" vector $\mathbf{X}^*$. In the Bayesian point of view, the notion of typicality discussed above is therefore very natural: we will describe the performance of inference procedures with high probability under the distributions of $\mathbf{X}^*$ and $\mathbf{Y}$, effectively discarding the very rare events that might influence a worst-case analysis.

Bayes-optimal setting and the Nishimori identity

An important concept that we will sometimes use to simplify the theoretical analysis is *Bayes-optimality*:

Definition 1.1 (*Bayes-optimal setting, informal*)

Let us consider an inference problem in which the observer wishes to recover a ground-truth vector $\mathbf{X}^* \in \mathbb{K}^n$ through a set of observations $\{Y_\mu\}_{\mu=1}^m \in \mathbb{R}^m$ which depend on $\mathbf{X}^*$. Such a problem is said to be *Bayes-optimal* if she/he knows the probabilistic models used to generate the vector $\mathbf{X}^*$ (called the *prior distribution*, $\mathbb{P}(\mathbf{x})$ in eq. (1.3)), and to generate the observations $\{Y_\mu\}$ given $\mathbf{X}^*$ (called the *channel distribution*, $\mathbb{P}(\mathbf{Y}|\mathbf{x})$ in eq. (1.3)).

Therefore, in the Bayes-optimal setting the observer knows the density of the posterior distribution $\mathbb{P}(\mathbf{x}|\mathbf{Y})$ by eq. (1.3)[2]. We will see that this knowledge allows to simplify a lot the theoretical analysis of estimation procedures. Although it may not seem a realistic assumption, we will see in the analyses of Part II that many results derived under a Bayes-optimality hypothesis transfer to more realistic setups in which the underlying distribution of the data is unknown.

In the Bayes-optimal setting, one can use a very important property called the *Nishimori identity*. Indeed, at fixed data samples $\mathbf{Y}$ we will often consider independent samples $\mathbf{x}_1, \cdots, \mathbf{x}_k$ drawn from the posterior distribution $\mathbb{P}(\mathbf{x}|\mathbf{Y})$ of eq. (1.3): such samples are called *replicas*. Physically speaking, the Nishimori identity shows that the planted solution $\mathbf{X}^\star$ behaves like another replica of the system.

Proposition 1.1 (*Nishimori identity*)

Let $(X, Y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ be a couple of random variables. Let $k \geq 1$ and let $X^{(1)}, \dots, X^{(k)}$ be k i.i.d. samples (given Y) from the conditional distribution $\mathbb{P}(X = \cdot\,|Y)$. Let us denote $\langle - \rangle$ the expectation operator w.r.t. $\mathbb{P}(X = \cdot\,|Y)$ and $\mathbb{E}$ the expectation w.r.t. (X, Y). Then, for all continuous bounded function g we have

$$\mathbb{E}\langle g(Y, X^{(1)}, \dots, X^{(k)})\rangle = \mathbb{E}\langle g(Y, X^{(1)}, \dots, X^{(k-1)}, X)\rangle\,.$$

Proof of Proposition 1.1 – This is a simple consequence of Bayes' formula. It is equivalent to sample the couple (X, Y) according to its joint distribution or to sample first Y according to its marginal distribution and then to sample X conditionally to Y from its conditional distribution $\mathbb{P}(X = \cdot\,|Y)$. Thus the $(k+1)$-tuple $(Y, X^{(1)}, \dots, X^{(k)})$ is equal in law to $(Y, X^{(1)}, \dots, X^{(k-1)}, X)$. This proves the proposition. □

Inference in high dimension

Modern algorithms which rule increasing parts of our lives, notably thanks to progress in deep learning, are mainly data-driven: they manage to learn a great number of parameters using comparably large numbers of data points. For instance, a classical neural network called GoogLeNet, introduced in 2015 for image classification [SLJ+15] and which achieved at the time state-of-the-art performance, possesses around 5 millions of internal parameters, and is trained on a dataset of more than one million images!

Our theory will focus on such inference procedures, in which the dimension of the internal parameters learned by algorithms is very large, while the number of data points available to the algorithm is also going to infinity, and both numbers are comparable. Anticipating on the connection with statistical physics, we call this setting the *thermodynamic limit*:

[2] Of course sampling from this high-dimensional distribution might still be very hard even when knowing its density.

Definition 1.2 (*Thermodynamic limit in inference models*)

Consider an inference (or learning) problem, in which one has to optimize over a set of parameters $\mathbf{x} \in \mathbb{K}^n$. In order to do so, we are given a number m of observations, or data samples in the context of learning. We will generically study a high-dimensional limit that we call *thermodynamic*, in which n and m both go to ∞, but their ratio remains finite $m/n \to \alpha > 0$.

1.1.3 Generalized Linear Models

Generalized Linear Models (GLMs) are a particularly important class of supervised learning models [NW72, McC18] that we will often take as example in the rest of Chapter 1, and that can be seen as broad generalizations of the perceptron algorithm of eq. (1.1). Their precise definition is the following:

Model 1.1 (*Generalized Linear Model (GLM)*)

Let $m, n \geq 1$. We are given a data, or sensing, matrix $\mathbf{F} \in \mathbb{K}^{m \times n}$. Given $\mathbf{F}$, data samples $\{Y_\mu\}$ are generated as:

$$\forall \mu \in \{1, \cdots, m\}, \qquad Y_\mu \sim P_{\text{out}}\Big(\cdot \Big| \frac{1}{\sqrt{n}} (\mathbf{F}\mathbf{X}^*)_\mu \Big),$$

in which $\mathbf{X}^* \in \mathbb{K}^n$ is the vector we will try to recover. $\mathbf{X}^\star$ is drawn with i.i.d. coordinates from a prior P_X, and P_{out} is a fixed probabilistic channel. *Compressive sensing* corresponds to a Gaussian channel distribution P_{out} with zero mean and variance $\Delta > 0$.

In this sense, GLMs generalize the usual linear regression by allowing the output function to be non-linear and possibly stochastic. They arise in many different areas of statistics, such as compressed sensing, phase retrieval [Fie82], or logistic regression. GLMs can also be thought of as the building blocks of fully-connected neural networks [LBH15], and we refer to [BKM+19] for a review of their numerous applications.

1.1.4 Gibbs-Boltzmann measure and the free entropy

Gibbs measure and posterior distribution – Let us briefly forget about inference and recall some basic elements of statistical mechanics. Consider a model of n particles x_i with i.i.d. distribution P_X, which interact via an energy function – or *Hamiltonian* – $\mathcal{H}(\mathbf{x})$ at inverse temperature $\eta = T^{-1} > 0$. The probability of a configuration $\mathbf{x}$ under the Gibbs distribution is proportional to the *Boltzmann weight*:

$$\mathbb{P}(\mathbf{x}) = \frac{1}{\mathcal{Z}_n} \exp\{-\eta \mathcal{H}(\mathbf{x})\} \prod_{i=1}^n P_X(x_i). \tag{1.4}$$

As we will see in a more mathematical way in Section 1.5, the Gibbs-Boltzmann measure described in eq. (1.4) arises naturally as the maximal-entropy distribution at a given temperature. The factor that ensures the proper normalization of the distribution is called the *partition function*, while its normalized logarithm is called the *free entropy*[3]:

$$\mathcal{Z}_n \equiv \int \mathrm{d}\mathbf{x} \prod_{i=1}^n P_X(x_i) \exp\{-\eta \mathcal{H}(\mathbf{x})\}, \qquad f_n \equiv \frac{1}{n} \ln \mathcal{Z}_n. \tag{1.5}$$

Very interestingly, all these notions are naturally transposed to the realm of inference models under a Bayesian point of view. To illustrate it, let us consider a GLM in the Bayes-optimal

[3] In this thesis we will mainly use this convention, while many physics works rather consider the negative of the free entropy, usually called *free energy*.

setting (cf. Definition 1.1 and Model 1.1), so that P_{out} and P_X are *known*. By Bayes' rule, the posterior distribution of $\mathbf{x}$ reads:

$$\mathbb{P}(\mathbf{x}|\mathbf{Y},\mathbf{F}) = \frac{\mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{Y}|\mathbf{x},\mathbf{F})}{\mathbb{P}(\mathbf{Y}|\mathbf{F})} = \frac{1}{\mathcal{Z}_n(\mathbf{Y},\mathbf{F})}\prod_{i=1}^{n} P_X(x_i)\prod_{\mu=1}^{m} P_{\text{out}}\Big[Y_\mu\Big|\frac{1}{\sqrt{n}}(\mathbf{F}\mathbf{x})_\mu\Big]. \tag{1.6}$$

It is clear that the distribution of eq. (1.6) can be easily mapped to the Boltzmann distribution of eq. (1.4), with $\eta = 1$ and energy given by:

$$\mathcal{H}(\mathbf{x}) \equiv -\sum_{\mu=1}^{m} \ln P_{\text{out}}\Big[Y_\mu\Big|\frac{1}{\sqrt{n}}(\mathbf{F}\mathbf{x})_\mu\Big]$$

The partition function is given by $\mathcal{Z}_n(\mathbf{Y},\mathbf{F}) = \mathbb{P}(\mathbf{Y}|\mathbf{F})$, and it will be instrumental in our asymptotic description of this problem. Note that in inference models the Gibbs distribution is a conditional distribution: it is parametrized by the observations $\mathbf{Y}$ (which implicitly contain the distribution of $\mathbf{X}^*$) and the input data (or sensing matrix) $\mathbf{F}$.

Notations – In general, we will denote $\langle\cdot\rangle$ the average with respect to the posterior distribution $\mathbb{P}(\mathbf{x}|\mathbf{Y})$, or the Gibbs distribution in statistical physics, while the symbol $\mathbb{E}$ will be used for average with respect to the other variables, e.g. $\mathbf{Y},\mathbf{F},\mathbf{X}^*$ in the GLM.

Free energy and mutual information – The average free entropy of an estimation model in the random setting is closely related to the *mutual information* between the observations and the signal, an important notion of information theory. Let us again illustrate it in the case of the GLM:

$$\begin{aligned}\frac{1}{n}I(\mathbf{X}^*;\mathbf{Y}|\mathbf{F}) &\equiv \frac{1}{n}\mathbb{E}_{\mathbf{Y},\mathbf{X}^*,\mathbf{F}}\ln\frac{\mathbb{P}(\mathbf{Y},\mathbf{X}^*|\mathbf{F})}{\mathbb{P}(\mathbf{X}^*)\mathbb{P}(\mathbf{Y}|\mathbf{F})} = \frac{1}{n}\mathbb{E}\ln\frac{\mathbb{P}(\mathbf{Y}|\mathbf{F},\mathbf{X}^*)}{\mathcal{Z}_n(\mathbf{Y},\mathbf{F})},\\ &= -\frac{1}{n}\mathbb{E}\ln\mathcal{Z}_n(\mathbf{Y},\mathbf{F}) + \frac{1}{n}\mathbb{E}\ln\mathbb{P}(\mathbf{Y}|\mathbf{F},\mathbf{X}^*).\end{aligned} \tag{1.7}$$

Therefore the mutual information and the free energy are equal up to an additive constant.

Energy-entropy competition – The Gibbs measure and the free entropy quantify the *competition between entropy and energy* in high-dimensional models, which is parametrized by the temperature parameter. Indeed, even if the minimizers of $\mathcal{H}(\mathbf{x})$ have very low energy, their mass under the "entropic" contribution to the measure (e.g. the term $\prod_i P_X(x_i)$ in eq. (1.4)) might be extremely small. In particular, at high temperature $\eta^{-1} \to \infty$, we expect for this reason the entropic contribution to dominate the physical state of the system. On the other hand, as $T = \eta^{-1} \downarrow 0$ the energetic contribution prevails, and the Gibbs measure concentrates on the actual minima of the Hamiltonian $\mathcal{H}(\mathbf{x})$.

1.1.5 Estimators

In order to gauge the performance of a procedure, we need a definition of the quantity we try to optimize. Let us describe two common estimators, again in the case of the GLM:

- **MMSE estimator** – The *Minimal Mean Squared Error* estimator $\hat{\mathbf{X}}_{\text{MMSE}}$ minimizes the L^2 distance between the estimator and the ground-truth signal. As we do not have access to $\mathbf{X}^*$, this distance is estimated using the posterior distribution, so that (recall that $\rho \equiv \mathbb{E}_{P_0}[X^2]$):

$$\hat{\mathbf{X}}_{\text{MMSE}} \equiv \arg\min_{\mathbf{x}} \frac{1}{n\rho}\mathbb{E}\int \mathrm{d}\mathbf{x}'\mathbb{P}(\mathbf{x}'|\mathbf{Y},\mathbf{F})\|\mathbf{x}-\mathbf{x}'\|^2. \tag{1.8}$$

As the MSE is a quadratic function, it is easy to see from eq. (1.8) that $\hat{\mathbf{X}}_{\text{MMSE}} = \mathbb{E}[\mathbf{x}|\mathbf{Y}, \mathbf{F}] = \langle \mathbf{x} \rangle$, that is the *marginal* under the posterior distribution. The achieved MMSE is:

$$\text{MMSE} = \frac{1}{n\rho}\mathbb{E}\|\mathbf{X}^* - \langle \mathbf{x} \rangle\|^2 = \rho - q, \tag{1.9}$$

with $q \equiv \mathbb{E}\{n^{-1}\sum_{i=1}^n X_i^* \langle x_i \rangle\}$. Eq. (1.9) follows from the Nishimori identity (Proposition 1.1) which gives $\mathbb{E}\{\|\langle \mathbf{x} \rangle\|^2\} = \mathbb{E}\{\mathbf{X}^* \cdot \langle \mathbf{x} \rangle\} = nq$.

- **MAP estimator** – The *Maximum A Posteriori* estimator maximizes directly the posterior distribution:

$$\hat{\mathbf{X}}_{\text{MAP}} \equiv \arg\max_{\mathbf{x}} \mathbb{P}(\mathbf{x}|\mathbf{Y}, \mathbf{F}). \tag{1.10}$$

In this thesis, we will usually favor the MMSE estimator over the MAP one. A first reason is that the MMSE provides directly a way to gauge the statistical significance of an estimate. Other reasons are the presence of overfitting in the MAP estimator in high dimensions, and the possibility that maximizing the posterior might be a very non-convex optimization problem, while the Bayes-optimal MMSE estimation is often easier to access with the tools we will develop. For more details and the introduction of other estimators in our framework we refer to [ZK16].

1.2 Intuitions from the physics of spin glasses

1.2.1 Why spin glasses?

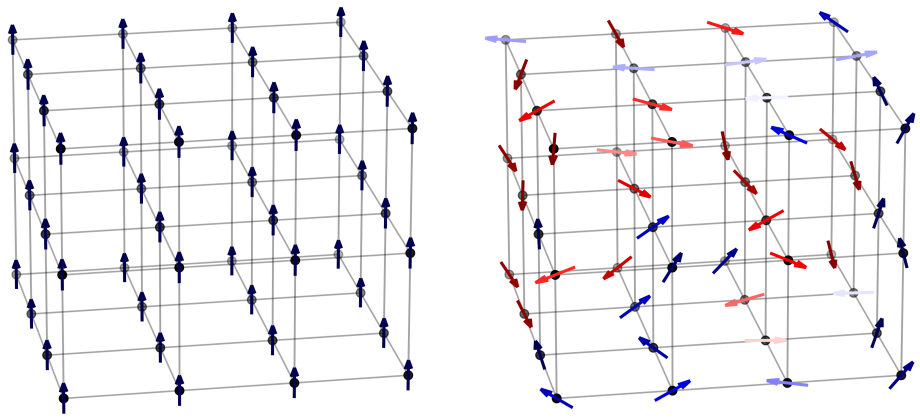
What is a spin glass ? – Spin glasses are a very peculiar class of physical states that were first observed in dilute magnetic alloys [Ste89]. While classical magnetic materials can be classified into two main categories known as *ferromagnetic* (when all magnetic moments align in the same direction) or *paramagnetic* (when all magnetic moments rapidly evolve with no global order), spin glasses are characterized by *frozen disorder*. In this phase the orientation of the magnetic spins shows no global order, but it evolves very slowly so that it is effectively frozen on short time scales. In 1975, Edwards and Anderson [EA75] introduced a statistical model of spins on the n-dimensional regular lattice which was found to exhibit such a glassy phase at low temperatures, paving the way for a theory of spin glasses. This model is very similar to the celebrated Ising model, and its interactions are described by the Hamiltonian:

$$\mathcal{H}_{\text{EA}}(\boldsymbol{\sigma}) = \sum_{(i,j)} J_{ij}\sigma_i\sigma_j, \quad \sigma_i = \pm 1 \tag{1.11}$$

where the sum runs over nearest-neighbors sites on a regular square lattice with n sites. In order to model the frozen disorder exhibited by spin glasses, Edwards and Anderson assumed that the interactions between sites were *random*, and took them to be Gaussian random variables $J_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(J_0/n, J^2/n)$. Depending on the values of J_0, J and the temperature T, they found that the system exhibited either paramagnetic, ferromagnetic, or glassy behavior, see Fig. 1.3.

It is important to stress that the randomness of the interactions is the critical feature that allows such models to exhibit glassy phases. In this dissertation, following a long line of theoretical work [MPV87], we thus usually call *spin glasses* statistical models with random disordered interactions.

Spin glasses and inference models – The correspondence between estimation problems and disordered systems in high dimensions is actually very natural. To illustrate it, let us consider



(A) A low-temperature ferromagnetic state in which all spins are aligned in a common direction.

(B) A *spin glass* state: the spins are frozen in random directions due to the randomness of the interactions.

FIGURE 1.3: Schematic representation of a ferromagnetic and a spin glass state in a 3-dimensional model with sites on a cube.

a flagship problem of statistics called *spiked matrix estimation*. The goal is to recover a signal vector $\mathbf{X}^*$ generated uniformly in $\{\pm 1\}^n$ from the observation of the "spiked" matrix

$$\mathbf{Y} \equiv \mathbf{W} + \sqrt{\lambda}\mathbf{X}^*(\mathbf{X}^*)^\intercal, \tag{1.12}$$

in which $\mathbf{W}$ is a random "noise" matrix, which we will consider here to be a standard Gaussian i.i.d. symmetric matrix, with $\mathbb{E}\, W_{ij}^2 = 1 + \delta_{ij}$. The posterior probability of $\mathbf{x}$ is

$$\begin{aligned}\mathbb{P}(\mathbf{x} = \boldsymbol{\sigma}|\mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y}|\mathbf{x} = \boldsymbol{\sigma})\mathbb{P}(\mathbf{x} = \boldsymbol{\sigma})}{\mathbb{P}(\mathbf{Y})} &= \frac{1}{2^n \mathbb{P}(\mathbf{Y})} \exp\Big\{ -\frac{1}{4}\mathrm{Tr}\Big[\Big(\mathbf{Y} - \sqrt{\lambda}\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\Big)^2\Big]\Big\}, \\ &\propto \exp\{\frac{\sqrt{\lambda}}{2}\sum_{i,j} Y_{ij}\sigma_i\sigma_j\}. \end{aligned} \tag{1.13}$$

The probability of eq. (1.13) is easily seen to be the Gibbs measure of a very simple disordered system at temperature $T = \lambda^{-1/2}$ with Hamiltonian:

$$\mathcal{H}_n(\boldsymbol{\sigma}) \equiv \frac{1}{2}\sum_{i,j} Y_{ij}\sigma_i\sigma_j.$$

As the interactions are given by the random matrix $\mathbf{Y}$, this Hamiltonian indeed describes a disordered model (note its similarity with eq. (1.11)). Therefore, studying the Gibbs measure of this long-range spin glass model with interaction matrix given by eq. (1.12) is equivalent to solving the spiked matrix estimation problem! More generally, many inference and learning tasks can be formulated as a statistical physics problem: the random interactions in spin glasses are mapped to the random quantities that parametrize the inference model (e.g. the noise matrix $\mathbf{W}$ in eq. (1.12) or the data matrix $\mathbf{F}$ in Model 1.1).

The wide extent of applications of spin glasses across the physical sciences was foreseen in [And89], an amazing two-page letter written in 1989 in which Anderson anticipated the pertinence of the rough high-dimensional landscapes of spin glasses in computer science and complexity theory. In fact, some of his interrogations on finding the ground state of the Sherrington-Kirkpatrick model (cf. Model 1.2) were only very recently answered [Mon21]. The general

connection of disordered systems with inference and optimization problems was first made explicit by Hopfield in [Hop82] for a simple model of neural networks, and by E. Gardner and coauthors for the perceptron algorithm [GD89]. Since then, it has been an extremely fruitful line of research (see e.g. [ZK16] for a review), and is an important guideline of this thesis. While the reader will find some basics of spin glass theory in this dissertation, she/he should refer e.g. to [MPV87] for more details, or to [CC05] for a classical introduction to spin glasses aimed at newcomers to the field.

1.2.2 Important concepts of spin glass theory

Annealed versus quenched, and self-averaging

An important concept in spin glass theory (and in high-dimensional probability) is the distinction between *annealed* and *quenched* averages[4]. It formalizes the difference between *typical* and *average* quantities for highly-fluctuating random variables. Let us illustrate this distinction first on a very simple example.

A toy example – Let X be a real random variable taking value e^{3n} with probability e^{-n}, and e^n with probability $1-e^{-n}$. Because $\lim_{n\to\infty}\mathbb{P}[X=e^n]=1$, the "typical" value of X is e^n. Since X is in the exponential scale, it is natural to rather consider the random variable $n^{-1}\ln X$. Its expectation is $n^{-1}\mathbb{E}\ln X = 2e^{-n}+1-e^{-n} = 1+\mathcal{O}_n(1)$: this average therefore describes well the typical behavior of X in the $n\to\infty$ limit, and it is trivial to check that the random variable $n^{-1}\ln X$ indeed concentrates on its mean. We call this limit the *quenched* average of X.

One could be tempted however to rather use $\mathbb{E}X$ to describe the typical behavior of X, as is common for non-fluctuating random variables. It is also exponentially large, and its leading order is given by $n^{-1}\ln\mathbb{E}X = n^{-1}\ln(e^{3n-n}+e^n-1) = 2+\mathcal{O}_n(1)$. This is what we call the *annealed* average of X, and it turns out here to be very different from the quenched one.

Where does this disparity come from? Actually, as the fluctuations of X are both very rare and very large, they dominate the naive average $\mathbb{E}X$! Annealed and quenched averages are here different because the annealed average is influenced by exponentially rare events to which the quenched average is immune.

Concentration: the blessing of dimensionality – In principle, the properties of a system depend on the realization of the disorder: e.g. the free entropy $n^{-1}\ln\mathcal{Z}_n(\mathbf{Y},\mathbf{F})$ of the GLM in eq. (1.6) depends on $\mathbf{Y},\mathbf{F}$. As this partition function $\mathcal{Z}_n$ is in general a highly-fluctuating random variable which scales exponentially in n, we define the *annealed and quenched free entropies* as:

$$\begin{cases} f_{\text{annealed}} &\equiv \lim\limits_{n\to\infty}\dfrac{1}{n}\ln\mathbb{E}\mathcal{Z}_n, \\ f_{\text{quenched}} &\equiv \lim\limits_{n\to\infty}\dfrac{1}{n}\mathbb{E}\ln\mathcal{Z}_n. \end{cases} \tag{1.14}$$

However, in high dimensions the fluctuations of intensive quantities often vanish, so that said quantities concentrate around their mean: this phenomenon is called *self-averaging* in statistical physics, and is particularly relevant for spin glasses. This is in particular the case of the free entropy $n^{-1}\ln\mathcal{Z}_n$ of many disordered systems, which is an intensive quantity that will self-average around its mean. This motivates our interest in f_{quenched} rather than f_{annealed}, as the first one characterizes the *typical* behavior of the system by the concentration of the free entropy:

$$\frac{1}{n}\ln\mathcal{Z}_n \xrightarrow[n\to\infty]{\mathbb{P}} f_{\text{quenched}}. \tag{1.15}$$

[4] This nomenclature comes from the name of metallurgic treatments.

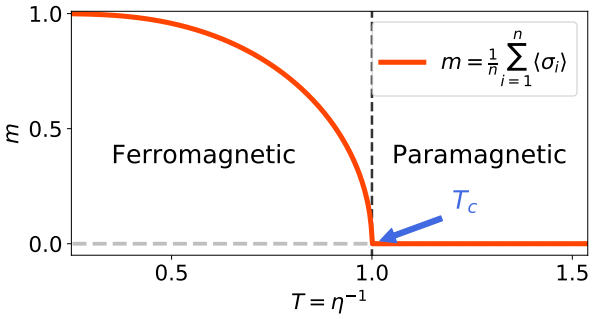


FIGURE 1.4: Average magnetization as a function of the temperature $T = \eta^{-1}$ for the Curie-Weiss model of eq. (1.17) with infinitely small external field $h \downarrow 0$.

Annealed approximation at high temperature – Note that by Jensen's inequality the annealed average of a random variable is always an upper bound of its quenched counterpart:

$$\mathbb{E} \ln X \leq \ln \mathbb{E} X. \tag{1.16}$$

At high temperature the partition function $\mathcal{Z}_n$ should not be dominated by rare events which are "smoothed out" by temperature, so that we expect the annealed average to be an informative approximation of the quenched free entropy in this regime. However it is very hard to control the accuracy of eq. (1.16) in general, and to gauge the significance of annealed results. Their main interest lies in the relative facility of their computation: as opposed to the quenched averages, they are often exactly computable by straightforward calculations.

"Quenched" variables – Besides the "quenched-annealed" terminology introduced above for asymptotic quantities, we will also call the variables on which we average in eq. (1.14) *quenched variables*: these are the "frozen" variables which parametrize the Gibbs distribution, e.g. the random interactions in a spin glass model. In the context of statistical estimation, the quenched variables will encompass several quantities, such as the planted solution $\mathbf{X}^\star$, the observations $\mathbf{Y}$, and the sensing matrix $\mathbf{F}$ in the case of the GLM, cf. Model 1.1.

Order parameters and phase transitions

The whole field of statistical physics is built around the idea that a complex system of infinitely many particles interacting with each other can be efficiently described on the macroscopic scale by a few simple quantities. Such quantities are called *order parameters*, and they can take many forms. For the sake of the presentation, let us focus on a non-disordered model called the Curie-Weiss model, at inverse temperature $\eta > 0$, and with a small external field $h > 0$ [Nis01]:

$$\mathbb{P}_\eta(\boldsymbol{\sigma}) \equiv \frac{1}{\mathcal{Z}_n(\eta)} \exp\Big\{ -\frac{\eta}{n} \sum_{i<j} \sigma_i \sigma_j - h \sum_{i=1}^{n} \sigma_i \Big\}. \tag{1.17}$$

In the Curie-Weiss model, as in non-disordered magnetic systems, the order parameter describing the macroscopic state of the system is the average *magnetization* $m \equiv n^{-1} \sum_{i=1}^{n} \langle \sigma_i \rangle$. Note that the Hamiltonian of eq. (1.17) favors *aligned spins*: the minimal energy is reached when all $\sigma_i = 1$. In Fig. 1.4 we show the evolution of m with the temperature $T = \eta^{-1}$, clearly describing the transition between a *paramagnetic* ($m = 0$) phase at high temperature, in which the entropic contribution to the free entropy dominates, and a *ferromagnetic* ($m > 0$, cf. Fig. 1.3a) phase, in which the spins are aligned to minimize the energy. The change between the two, called a *phase transition*, happens at the critical temperature $T_c = 1$.

In systems with a random disorder, such as spin glasses, the magnetization is however not a good order parameter in general. This can be intuitively understood by Fig. 1.3b: while all the local spins are "frozen" in specific directions – due to the randomness of the interactions – there is no global orientation of the spins, and the total magnetization will always be equal to zero. In this class of models, one must turn to another order parameter: the *overlap*[5], also called the *Edwards-Anderson* order parameter [EA75]:

$$q_{\mathrm{EA}} \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\langle\sigma_i\rangle^2]. \tag{1.18}$$

The overlap will be non-zero as soon as all local magnetizations are non-zero, even if they do not share a global direction. It will be instrumental in describing phase transitions arising in spin glasses and inference models in this thesis.

1.2.3 Some classical spin glass models

To conclude this rapid presentation of spin glasses let us define three important models that will be of interest in several parts in this manuscript. In order to define the first one we anticipate a bit on our introduction of random matrices, and describe a first random matrix ensemble known as the Gaussian Orthogonal (or Unitary) Ensemble.

Definition 1.3 (*Gaussian Orthogonal/Unitary Ensemble*)

We say that a matrix $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ is drawn from the *Gaussian Orthogonal Ensemble* (if $\mathbb{K} = \mathbb{R}$) or *Gaussian Unitary Ensemble* (if $\mathbb{K} = \mathbb{C}$), and we write $\mathbf{J} \sim \mathrm{GOE}(n)$ (resp. $\mathbf{J} \sim \mathrm{GUE}(n)$), if we have $J_{ij} \sim \mathcal{N}_\beta(0, (1+\delta_{ij})/n)$ for $i \leq j$, and all $\{J_{ij}\}_{i\leq j}$ are independent.

This allows to define a spin glass model, introduced by Sherrington and Kirkpatrick in [SK75], which played a cardinal role in the history of the field.

Model 1.2 (*Sherrington-Kirkpatrick*)

Let $n \geq 1$ and $\mathbf{J} \sim \mathrm{GOE}(n)$. The *Sherrington-Kirkpatrick* Hamiltonian is defined as:

$$H_{n,\mathbf{J}}(\boldsymbol{\sigma}) \equiv \sum_{i<j} J_{ij}\sigma_i\sigma_j, \qquad \boldsymbol{\sigma} \in \{\pm 1\}^n. \tag{1.19}$$

Another important class of spin glasses that has received tremendous attention since the 1990s is the class of *p-spin models*. The simplest instance of this class is known as the *pure* p-spin, and we define it here on the high-dimensional sphere.

Model 1.3 (*Pure spherical p-spin*)

Let $n \geq 1$ and $p \geq 2$. The *pure p-spin* model Hamiltonian is defined as:

$$H_{n,p}(\boldsymbol{\sigma}) \equiv \frac{1}{n^{\frac{p-1}{2}}} \sum_{i_1,\cdots,i_p} J_{i_1,\cdots,i_p}\sigma_{i_1}\cdots\sigma_{i_p}, \qquad \boldsymbol{\sigma} \in \mathbb{S}^{n-1}(\sqrt{n}), \tag{1.20}$$

with $J_{i_1,\cdots,i_p} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. $H_{n,p}$ is therefore a Gaussian random field with zero mean and covariance $\mathbb{E}[H_{n,p}(\boldsymbol{\sigma})H_{n,p}(\boldsymbol{\sigma}')] = n(\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'/n)^p$.

In mathematical terms, the spherical pure p-spin corresponds to the simplest example of a random function: a homogeneous polynomial of degree p with Gaussian coefficients. We can generalize the model further, to what is known as the *mixed p-spin*.

[5] It is called *overlap* as it corresponds to the scalar product of the magnetization vectors of two independent "replicas" of the system which share the same quenched disorder.

Model 1.4 (*Mixed spherical p-spin*)

Let $n \geq 1$ and a real sequence $(c_p)_{p\geq 2}$ with $\sum_{p\geq 2} c_p^2 < \infty$. The *mixed p-spin* model Hamiltonian is defined as:

$$H_n(\boldsymbol{\sigma}) \equiv \sum_{p=2}^{\infty} c_p H_{n,p}(\boldsymbol{\sigma}), \qquad \boldsymbol{\sigma} \in \mathbb{S}^{n-1}(\sqrt{n}). \tag{1.21}$$

Equivalently, H_n is a Gaussian random field with zero mean and covariance $\mathbb{E}[H_n(\boldsymbol{\sigma})H_n(\boldsymbol{\sigma}')] = n\xi(\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'/n)$, with the *mixture* function $\xi(t) \equiv \sum_{p=2}^{\infty} c_p^2 t^p$.

By a classical result of Schoenberg [Sch42], the mixed p-spin models span all stationary isotropic Gaussian random fields on the sphere $\mathbb{S}^{n-1}(\sqrt{n})$. Note that Models 1.3 and 1.4 can easily be generalized to a variety of prior distributions different from the spherical one, e.g. one can consider Ising spins on the hypercube $\boldsymbol{\sigma} \in \{\pm 1\}^n$, as in the SK model 1.2.

1.3 Static approximations to the free energy

1.3.1 Replica theory and replica symmetry breaking

As we emphasized in Section 1.2, one of our most important tasks will be to compute *quenched* averages, which (as opposed to annealed ones) are representative of the typical behavior of a high-dimensional system. Unfortunately, quenched quantities are in general much harder to compute than their annealed counterparts, and the *replica method* was developed precisely to tackle this difficulty. Its first application to disordered systems goes back to Edwards & Anderson [EA75], and since then it has achieved tremendous success in the study of spin glasses, but also of inference problems. More precise introductions to the beautiful field of replica theory can be found in [MPV87, Nis01, CC05, MM09].

For the sake of the presentation, let us focus on the Sherrington-Kirkpatrick Hamiltonian (Model 1.2). In the context of inference models the calculations are similar, and we refer the reader to Appendix B which precisely details two replica computations in high-dimensional estimation. As we detailed above, in order to characterize the typical behavior of the model, our goal will be to compute the quenched free energy at inverse temperature $\eta > 0$:

$$\Phi(\eta) \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{\mathbf{J}} \ln \mathcal{Z}_n, \text{ with } \mathcal{Z}_n \equiv \sum_{\mathbf{x}\in\{\pm 1\}^n} \exp\Big\{ -\frac{\eta}{\sqrt{n}} \sum_{i,j} J_{ij} x_i x_j \Big\}. \tag{1.22}$$

Note that we rescaled the interaction variables J_{ij} by a factor $\sqrt{n}$ so that $\mathbb{E}J_{ij}^2 = 1 + \delta_{ij}$.

The replica trick to compute quenched averages

The difficulty in computing the quenched average in eq. (1.22) is the expectation of the *logarithm* of a highly-fluctuating quantity. In [EA75], Edwards & Anderson proposed an informal way, now known as the *replica trick*, to compute such averages. It is based on the following identity:

$$\mathbb{E}_{\mathbf{J}} \ln \mathcal{Z}_n = \lim_{p\downarrow 0} \frac{1}{p} \ln \mathbb{E}_{\mathbf{J}}[\mathcal{Z}_n^p]. \tag{1.23}$$

While such an identity is a priori correct, the hope of the replica method is to compute the RHS of eq. (1.23) for *integer* p (i.e. the moments of $\mathcal{Z}_n$) and to analytically expand this expression for arbitrary $p > 0$, before taking the limit $p \downarrow 0$. The actual replica method relies on an additional

heuristic inversion of the $p \downarrow 0$ and $n \to \infty$ limits, so that we will refine eq. (1.23) by

$$\Phi(\eta) \equiv \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\mathbf{J}} \ln \mathcal{Z}_n = \lim_{p\downarrow 0} \lim_{n\to\infty} \frac{1}{np} \ln \mathbb{E}_{\mathbf{J}}[\mathcal{Z}_n^p]. \tag{1.24}$$

Let us apply this identity to the SK model. We obtain by direct Gaussian integration:

$$\begin{aligned} \frac{1}{n} \ln \mathbb{E}_{\mathbf{J}}[\mathcal{Z}_n^p] &= \frac{1}{n} \ln \int \prod_{i\leq j} \mathrm{d}J_{ij} \frac{e^{-\frac{1}{4}\sum_{i,j} J_{ij}^2}}{(2\pi)^{\frac{n(n-1)}{4}} (4\pi)^{n/2}} \prod_{a=1}^{p} \sum_{\mathbf{x}^a \in \{\pm 1\}^n} \exp\Big\{ -\frac{\eta}{2\sqrt{n}} \sum_{a=1}^{p} \sum_{i,j} J_{ij} x_i^a x_j^a \Big\}, \\ &= \frac{1}{n} \ln \prod_{a=1}^{p} \sum_{\mathbf{x}^a \in \{\pm 1\}^n} \exp\Big\{ n \frac{\eta^2}{4} \sum_{1\leq a,b\leq p} \Big(\frac{1}{n} \sum_{i=1}^{n} x_i^a x_i^b \Big)^2 \Big\}. \end{aligned} \tag{1.25}$$

Let us denote $\mathbf{Q} \in \mathcal{S}_p(\mathbb{R})$ the *overlap matrix* that appears in eq. (1.25):

$$Q^{ab} \equiv \frac{1}{n} \sum_{i=1}^{n} x_i^a x_i^b. \tag{1.26}$$

Note in particular that $Q^{aa} = 1$. Introducing $\mathbf{Q}$ in eq. (1.25) and using the exponential representation of the delta function we reach:

$$\begin{aligned} &\frac{1}{n} \ln \mathbb{E}_{\mathbf{J}}[\mathcal{Z}_n^p] = \\ &\frac{1}{n} \ln \int \prod_{1\leq a,b\leq p} \mathrm{d}Q^{ab} \mathrm{d}\hat{Q}^{ab} e^{\frac{n}{2}\sum_{a,b}\left[\frac{\eta^2}{2}(Q^{ab})^2 + Q^{ab}\hat{Q}^{ab}\right]} \left[\sum_{x^1,\cdots,x^p=\pm 1} e^{-\frac{1}{2}\sum_{a,b} \hat{Q}^{ab} x^a x^b} \right]^n + \mathcal{O}_n(1). \end{aligned} \tag{1.27}$$

Replica symmetric assumption

When looking at eq. (1.27), one can make wo important remarks:

- In eq. (1.27) we can apply, as $n \to \infty$, Laplace's method on the variables $\{Q^{ab}, \hat{Q}^{ab}\}$. Indeed, the number of these variables is $\mathcal{O}_n(1)$.
- The application of Laplace's method tells us that the overlaps Q^{ab} are critical in characterizing the thermodynamical state of the system. For this reason these variables are called *order parameters*, see Section 1.2.2.

Nevertheless, Laplace's principle in eq. (1.27) is quite hard to formulate. Recall indeed that, in order to apply the replica method, we need the LHS of eq. (1.27) as an analytical function of p! This leads us to rely on physical arguments to deduce the correct form of $\mathbf{Q}, \hat{\mathbf{Q}}$ at the solution of Laplace's method. When confronted with this problem the first solution that comes to mind is very natural: as all replicas should be equivalent, the matrix $\mathbf{Q}$ at the solution should have a *replica-symmetric* form

$$Q^{ab} = q_0 + (1 - q_0)\delta_{ab}. \tag{1.28}$$

We assume a similar form of $\hat{Q}^{ab} = -\hat{q}_0(1 - \delta_{ab})$. This was the assumption of Sherrington & Kirkpatrick [SK75] who proposed a first solution to this model. After some straightforward calculations[6] we obtain from eq. (1.27) an analytic expression in p:

$$\frac{1}{n} \ln \mathbb{E}_{\mathbf{J}}[\mathcal{Z}_n^p] = \frac{\eta^2 p}{4} + \eta^2 \frac{p(p-1)}{4} q_0^2 - \frac{p(p-1)}{4} q_0 \hat{q}_0 - \frac{p\hat{q}_0}{2} + \int \mathcal{D}\xi \{2\cosh(\sqrt{\hat{q}_0}\xi)\}^p + \mathcal{O}_n(1).$$

[6] Using in particular the identity $e^{x^2/2} = \int \mathcal{D}\xi\, e^{\xi x}$.

Recall that $\mathcal{D}\xi$ is the standard Gaussian law $\mathcal{N}(0,1)$. Applying then the replica method in the form of eq. (1.24) we reach finally the replica-symmetric free entropy for the SK model:

$$\Phi = \frac{\eta^2(1-q_0^2)}{4} + \frac{q_0\hat{q}_0}{4} - \frac{\hat{q}_0}{2} + \int \mathcal{D}\xi \ln\left\{2\cosh(\sqrt{\hat{q}_0}\xi)\right\}. \tag{1.29}$$

In this equation one should extremize with respect to the set of variables $(q_0, \hat{q}_0)$. We can make several remarks from eq. (1.29) (details can be found in [SK75] or [Nis01]):

- There exists a transition point $\eta_c = 1$. For $\eta \leq \eta_c$, the solution that maximizes eq. (1.29) is $q_0 = 0$, while $q_0 > 0$ for $\eta > \eta_c$. Recall that from eq. (1.26) one can easily see that $q_0 = n^{-1}\sum_i \langle x_i \rangle^2$ is the Edwards-Anderson order parameter. This implies that at $T_c = \eta_c^{-1} = 1$ there is a transition from a *paramagnetic* phase (in which $q_0 = 0$) to a *spin glass* phase (in which $q_0 > 0$). Moreover, in both phases the average magnetization $m \equiv n^{-1}\sum_i \langle x_i \rangle = 0$.
- However very puzzling behaviors happen when $T = \eta^{-1}$ approaches zero. Let us cite [SK75]: "The entropy S [...] goes to a negative limit at $T = 0$. We speculate that this unphysical behavior has its origin in the interchange of limits $n \to \infty$ and $p \downarrow 0$, but that the consequences are confined to low temperatures."

This second remark implies that one of our assumptions breaks down at sufficiently low temperatures. However the intuition of [SK75] was not really exact: what happens actually is that at $T < T_c$ the replica-symmetric solution of eq. (1.28) is not stable in overlap space when applying Laplace's method in eq. (1.27), so that the actual maximum is reached in a point that breaks the symmetry between the replicas...

A word on replica symmetry breaking

In a series of beautiful papers [Par79, Par80a, Par80b], Giorgio Parisi proposed with a spark of genius the solution to the issue mentioned above. What he understood was that the overlap q between two replicas of the system with the same interactions, i.e. the order parameter of the system, could have a very non-trivial behavior when $n \to \infty$. More precisely, its probability distribution $P(q)$ is in general involved, and can not be characterized by a single q_0 as we assumed in the replica-symmetric ansatz. Without diving into details (which the reader will find e.g. in [CC05] or [MPV87]), one brilliant realization was the mapping of a generic $P(q)$ to a point in replica space, i.e. a Q^{ab} that generalized the replica-symmetric form of eq. (1.28).

First of all, note that the replica-symmetric ansatz corresponds to a distribution $P(q) = \delta(q - q_0)$: in probabilistic terms, replica symmetry is equivalent to the concentration (or self-averaging) of the overlap, an important intuition to keep in mind at several points of this thesis. In the SK model detailed above the overlap is not concentrating for $T < T_c$, and the RS solution has to be discarded[7]. The next step is to consider a $P(q)$ made of two delta peaks. One can show that in the limit $p \downarrow 0$, it corresponds to the following overlap matrix:

$$P(q) = x\delta(q-q_0) + (1-x)\delta(q-q_1) \Longleftrightarrow \mathbf{Q} = \begin{pmatrix} 1 & q_1 & q_1 & & & & \\ q_1 & 1 & q_1 & & \cdots & q_0 & \cdots \\ q_1 & q_1 & 1 & & & & \\ & & & \ddots & & & \\ & & & & 1 & q_1 & q_1 \\ \cdots & q_0 & \cdots & & q_1 & 1 & q_1 \\ & & & & q_1 & q_1 & 1 \end{pmatrix}. \tag{1.30}$$

[7] Note however that it allows to predict the correct paramagnetic - spin glass transition at $T_c = \eta_c = 1$.

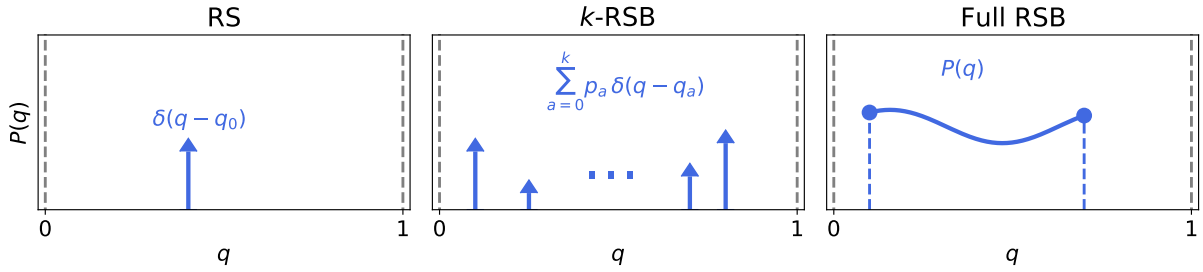


FIGURE 1.5: The different forms of the Parisi measure $P(q)$ corresponding to replica symmetry (left), k-th level of replica symmetry breaking (center), and full replica symmetry breaking (right).

The inner blocks with off-diagonal term q_1 have size $x \in \{1, \cdots, p\}$ (here we show $x = 3$). However (another oddness of the replica method which contradicts all usual mathematical intuitions) since we eventually take the limit $p \downarrow 0$, the parameter x will lie in the interval $[0, 1]$ so that the parameterization of $P(q)$ above is well-defined. Going further, one can iterate the hierarchical structure of the overlap matrix in eq. (1.30) to create an arbitrary number of delta peaks in $P(q)$, called the *replica symmetry breaking (RSB)* level (e.g. the distribution of eq. (1.30) is 1-RSB). When $k \to \infty$, we say that the system is in a *full replica symmetry breaking (FRSB)* phase, and the support of $P(q)$ is continuous. We illustrate the different possible shapes of $P(q)$ depending on the level of RSB in Fig. 1.5. In the SK model we considered above, it was shown by Parisi [Par79] that for $\eta \geq \eta_c$ the system is in a FRSB phase, and that the overlap distribution $P(q)$ (also called the *Parisi measure*) has a continuous support $[0, q^*(\eta)]$. The FRSB picture solved the negative-entropy problem, and after decades of mathematical works the Parisi solution for the SK model was eventually proven by Talagrand [Tal06], putting on firm ground one of the most important predictions of the physics of disordered systems.

RSB and the form of the Gibbs measure – Roughly speaking, 1-RSB corresponds to a space of local minima of the free energy which is organized into exponentially many clusters. Inside each cluster two solutions typically have overlap q_1, while solutions belonging to two different clusters have a typical overlap q_0. This hierarchy can be iterated inside each cluster, which gives rise to the 2-RSB structure. Iterating even further, the level of RSB corresponds to the depth of this hierarchical structure of clusters, which is known as an *ultrametric structure*. Ultrametricity and RSB is an incredibly beautiful mathematical representation of the free energy landscape of spin glass models, which also allows to create efficient algorithms, something which was only understood very recently [AMS20, AM20, Sub21, Mon21]. Unfortunately, a precise description would be beyond the scope of this dissertation (and would surely require an additional chapter), and we refer instead to [MPV87] for more details on this subject.

Final remarks

Our discussion of replica theory was very brief, as our main goal was to introduce the ideas behind the method, without diving too much into its physical consequences, which are extensively discussed in the literature mentioned above. To conclude this part, as we focused a lot on spin glasses, let us make two remarks on the application to estimation models.

Application to inference problems – In high-dimensional Bayesian inference (e.g. in the Generalized Linear Model, cf. Section 1.1), the role of the quenched interactions J_{ij} is played by the *data samples* (the matrix $\mathbf{F}$ in the GLM). Leveraging on this analogy we can apply the replica method to these models to compute the *quenched* free entropy, characteristic of typical realizations of the data. As in spin glasses, the order parameter governing the state of the system in the high-dimensional limit is the *overlap* q between two replicas, to which we must add the overlap m between a replica and the planted solution $\mathbf{X}^\star$. As we saw in

Section 1.1 this overlap is related to the asymptotic MMSE. Historically, the first applications of the replica method to the optimal performance in inference models date back to studies of the perceptron [GD89, Gyö90, STS90]. The reader interested in more concrete applications of the replica method to inference problems should refer to Appendix B in which we detail two replica calculations, in a model of neural networks and in phase retrieval.

No replica symmetry breaking in Bayes-optimal problems – An important property that we will use several times throughout this thesis is that in Bayes-optimal inference (see Section 1.1), replica symmetry is never broken. This was shown in several ways and can intuitively be understood as follows. By Bayes-optimality the planted signal $\mathbf{X}^*$ can just be seen as an additional replica (that we will index as $\mathbf{x}^0 = \mathbf{X}^\star$), which, combined with the Nishimori identity, implies that the couple of replicas $(0, a)$ and (a, b) should be equivalent for all a, b, i.e. the system is replica-symmetric. For more details on this argument, we refer the reader to [ZK16].

1.3.2 Thouless-Anderson-Palmer approach

Pure states and TAP free energy

We present here a derivation, first performed by Thouless, Anderson and Palmer [TAP77] for the SK model, of what is now known as the TAP free energy. Note that the replica method described in Section 1.3.1 only allows to access the average free energy. However at finite temperature we would also like to access properties of a specific instance of the system: we will achieve this feat by characterizing the position of what is known as *pure states*.

Indeed, Thouless Anderson and Palmer [TAP77] understood that at a given $\eta \geq 0$ the Gibbs measure typically concentrates all its mass on a large number of small regions of the landscape. These regions are characterized by their barycenter (also called magnetization) $\mathbf{m} \in \mathbb{R}^n$, and they are known as pure states. When the temperature approaches zero, these pure states approach the global minima of the original Hamiltonian. To summarize, we can decompose the mean of every observable O at inverse temperature η as:

$$\langle O \rangle = \sum_\alpha w_\alpha \langle O \rangle_\alpha, \tag{1.31}$$

in which $\langle \cdot \rangle_\alpha$ is the average over a single pure state, and w_α are the weights of each pure state under the Gibbs measure. We illustrate this in an informal way in Fig. 1.6. Note that this description has very recently partially been put on rigorous ground for spin glass models in the mathematics literature [FMM21, CPS21].

We wish to build a free entropy corresponding to the decomposition into pure states, focusing on the SK model for introducing the method. Stated differently, what we want is a function of the local magnetization $\mathbf{m}$, whose local maxima are the pure states of the system. In order to achieve this, we will tilt the original Gibbs measure of the system (see eq. (1.22)) in order to constraint $\langle x_i \rangle = m_i$. The idea is that the free entropy associated to this constrained Gibbs measure will be maximal when $\mathbf{m}$ belongs to the pure states on which the Gibbs measure decomposes. This function is what we call the *Thouless-Anderson-Palmer (TAP)* free entropy. Introducing Lagrange multipliers $\{\lambda_i\}$ to enforce the constraints $\langle x_i \rangle = m_i$, it reads:

$$\Phi_{\text{TAP}}(\eta, \mathbf{m}) \equiv \frac{1}{n}\sum_{i=1}^n \lambda_i m_i + \frac{1}{n} \ln \sum_{\mathbf{x}\in\{\pm 1\}^n} \exp\Big\{ -\sum_{i=1}^n \lambda_i x_i - \frac{\eta}{2\sqrt{n}} \sum_{i,j} J_{ij} x_i x_j \Big\}, \tag{1.32}$$

in which we implicitly extremize over the $\{\lambda_i\}$. In particular, when $\mathbf{m}$ is a global maximum of Φ_{TAP}, $\Phi_{\text{TAP}}(\mathbf{m})$ should be the actual free entropy of the system, that we can compute e.g. with the replica method, cf. Section 1.3.1. The behavior of the TAP free energy landscape gave

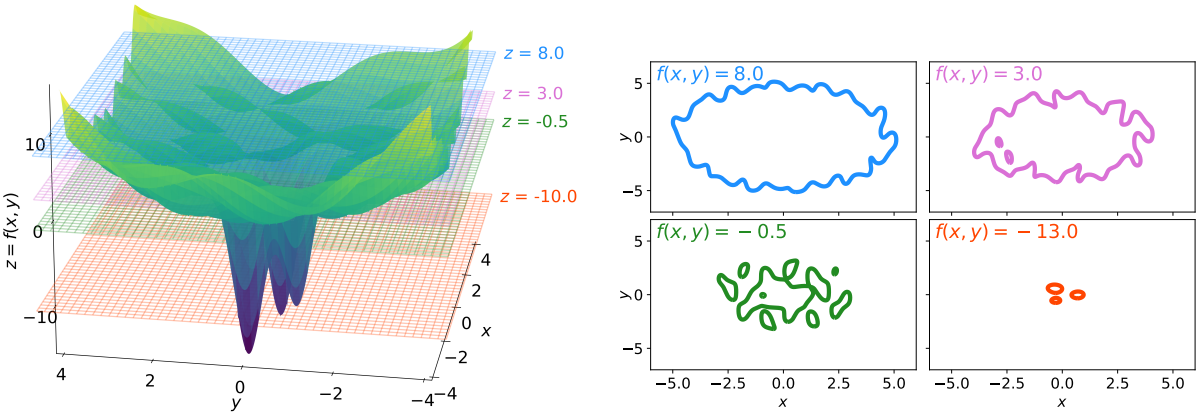


FIGURE 1.6: (Left) A generic "rough" landscape $f(x, y)$ with several deep minima in dimension 2. (Right) Slices $z = f(x, y)$, shown as planes in the left picture. For z high enough (high energy, or high temperature by the microcanonical-canonical equivalence in the thermodynamic limit) the slices are well-defined by a single "pure state". As the energy decreases the slices get disconnected, and for low enough energy they are concentrated around the location of the deep minima of $f(x, y)$. This low-dimensional view is merely an intuition and should not be considered as anything more.

rise to a rich literature, both from the physics and mathematics perspective, and we refer to Chapter 2 for more details on the TAP approach.

The TAP free entropy of eq. (1.32) can be computed in different ways: the perhaps simplest route is through a low-η expansion, introduced by Plefka [Ple82] and later made systematic in [GY91]. We will make extensive use of these expansions in Chapters 2 and 3, so that we do not detail their application to the SK model and leave the derivation of the following equation to the reader:

$$\Phi_{\mathrm{TAP}}(\eta, \mathbf{m}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1+m_i}{2}\ln\frac{1+m_i}{2} + \frac{1-m_i}{2}\ln\frac{1-m_i}{2}\right\} + \frac{\eta}{2n^{3/2}}\sum_{i,j}J_{ij}m_i m_j + \frac{\eta^2}{4n^2}\sum_{i,j}J_{ij}^2(1-m_i^2)(1-m_j^2) + \mathcal{O}_n(1). \tag{1.33}$$

Let us briefly discuss the different terms of eq. (1.33) and their physical consequences.

First order: naive mean-field

Truncating the series of eq. (1.33) at order 1 in η yields what is known as the *mean-field* approximation:

$$\Phi_{\mathrm{MF}}(\eta, \mathbf{m}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1+m_i}{2}\ln\frac{1+m_i}{2} + \frac{1-m_i}{2}\ln\frac{1-m_i}{2}\right\} + \frac{\eta}{2n^{3/2}}\sum_{i,j}J_{ij}m_i m_j. \tag{1.34}$$

It is instructive to consider the maximization equations of this mean-field free entropy. They read $m_i = \tanh\left(\eta\sum_{j=1}^{n}J_{ij}m_j/\sqrt{n}\right)$, that is $\langle x_i\rangle = \tanh\left(\eta\sum_{j=1}^{n}J_{ij}\langle x_j\rangle/\sqrt{n}\right)$. However, the actual exact equations that one can easily derive from eq. (1.32) read $\langle x_i\rangle = \left\langle\tanh\left(\eta\sum_{j=1}^{n}J_{ij}x_j/\sqrt{n}\right)\right\rangle$. We therefore see quite clearly the approximation: it neglects the fluctuations of the effective field $\sum_j J_{ij}x_j$ felt by the spin x_i, and replaces it with its mean (hence the name "mean-field").

Mean-field as a variational approximation – The mean-field approximation can be recovered by a different avenue, namely by a variational principle. Indeed, let us denote μ_η the Gibbs measure at inverse temperature $\eta \geq 0$. Then minimizing the Kullback-Leibler divergence

$D_{\mathrm{KL}}(\nu|\mu_\eta)$ can be tractable provided that we minimize it on an appropriate set of measures. The mean-field approximation neglects correlations between variables, and therefore corresponds to minimizing over the set of factorized measures, parametrized by their mean m_i:

$$\frac{\mathrm{d}\nu}{\mathrm{d}x} = \prod_{i=1}^{n} \Big\{ m_i \delta(x-1) + (1-m_i)\delta(x+1) \Big\}.$$

For the detailed variational derivation of the mean-field approximation, and how it yields back eq. (1.34), one can refer to [Gab20].

Higher orders and Onsager correction

The naive mean-field approximation can be exact provided the interactions are long-ranged and weak enough, e.g. in the Curie-Weiss model. While the SK model 1.2 is long-ranged, its interactions are not weak enough, so that the actual TAP free entropy of eq. (1.33) contains an additional term with respect to the mean-field, which is of order η^2 and is called the *Onsager reaction term*. Let us write the complete maximization equations of eq. (1.33) (called the *TAP equations*), that one has to solve in order to find the pure states of the system. For use in a further argument, we also add an external field $\sum_i h_i m_i$ to the Hamiltonian, which does not affect the free entropy of eq. (1.33) beyond the mean-field term. In the end, the TAP equations read:

$$m_i = \tanh\left(\frac{\eta}{\sqrt{n}} \sum_{j=1}^{n} J_{ij} m_j + \eta h_i - \left[\frac{\eta^2}{n} \sum_{j=1}^{n} J_{ij}^2 m_j^2 \right] m_i \right). \tag{1.35}$$

The last term in the RHS of eq. (1.35) is a reaction term: indeed, we saw that the site i affects its neighboring sites j by an effective field $J_{ij}m_j$. This effective field therefore modifies the value of m_j by $\delta m_j = \chi_{jj} J_{ij} m_i$, in which $\chi_{jj} \equiv \partial m_j / \partial h_j$ is the susceptibility of the site j. By eq. (1.35) we have $\chi_{jj} = \eta(1-m_j^2)$, so that m_j is modified by an amount $\delta m_j = \eta J_{ij} m_i (1-m_j^2)$. The Onsager reaction term then arises to cancel the retro-action of this modified value of m_j on m_i. Indeed, by our arguments, the field sent by δm_j in the direction of m_i is equal to $J_{ij}\delta m_j = \eta J_{ij}^2 (1-m_j)^2 m_i$. Summing these corrections over all the neighbors of i yields the Onsager reaction term of eq. (1.35). Roughly speaking, this term amounts to remove the effect of the retro-action of m_i on itself, which was wrongfully taken into account in the mean-field approximation.

Orders $k \geq 3$ – Note that in the TAP free entropy of the SK model, cf. eq. (1.33), there is no contributions of terms of order η^k with $k \geq 3$. However, in spin glass or inference models with non-Gaussian interactions one needs in general to compute the whole perturbation series at any order in η, which complicates drastically the reaction terms. We will analyze such models in Chapter 2, and in particular we will show how to compute these reaction terms using the free cumulants of the interaction matrix, an important object of random matrix theory defined in the following Section 1.5.

1.4 From physics to algorithms

In Section 1.3 we focused on different approaches to compute the free energy (or free entropy), or to characterize the pure states of disordered systems. This leaves unanswered many important questions: if we are given a practical instance of the problem, are there efficient ways to find an approximate global minimum of the Hamiltonian? Can we sample in polynomial time from the Gibbs-Boltzmann distribution? Can we compute the asymptotic free entropy? All these considerations fall within the scope of *algorithmic* studies.

In this part we present a class of algorithms, known under the umbrella of *message-passing*, and which will allow us to study optimal performances in a variety of high-dimensional estimation problems.

A note on gradient-descent algorithms – While they will not be investigated directly here, let us mention another popular class of procedures that are also related to the physics point of view. The perhaps most natural approach to design an algorithm minimizing an energy function is to consider a local gradient-descent algorithm (or one of its refinements – e.g. stochastic gradient descent), a strategy which is naturally transposed to inference problems. For instance, in supervised learning a classical procedure is *empirical risk minimization* (ERM): we wish to learn a predictor function $h(x) : \mathcal{X} \to \mathcal{Y}$, and we are given a set of input-output data points $\{x_\mu, y_\mu\}_{\mu=1}^m \in (\mathcal{X} \times \mathcal{Y})^m$. In ERM one chooses a loss function $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ and minimizes the empirical risk over the given space of predictors h:

$$\hat{\mathcal{R}}_m(h, \{x_\mu, y_\mu\}) \equiv \frac{1}{m} \sum_{\mu=1}^{m} l(y_\mu, h(x_\mu)). \tag{1.36}$$

Empirical risk minimization with local gradient-descent algorithms has known tremendous success in the past decade, so that these methods are now widely used in the field of machine learning [Bot03], and they are a major ingredient of the current success of deep neural networks [Zde20]. From the physics point of view, such methods are related to the celebrated *Langevin dynamics*, whose performances have been analyzed for spherical spin glasses [CK93, BADG06]. This analysis has recently been generalized to the context of inference, for a class of spiked matrix-tensor spherical models [SMBC$^+$19, SMKUZ19, SMBC$^+$20b], and the construction of a theoretical framework for tracking the performance of gradient-based algorithms in learning is a very dynamic line of research [GAS$^+$20, SMBC$^+$20a, MKUZ20, MUZ21]. Finally, gradient-based optimization will motivate the topological analysis undertook in Part III of this dissertation.

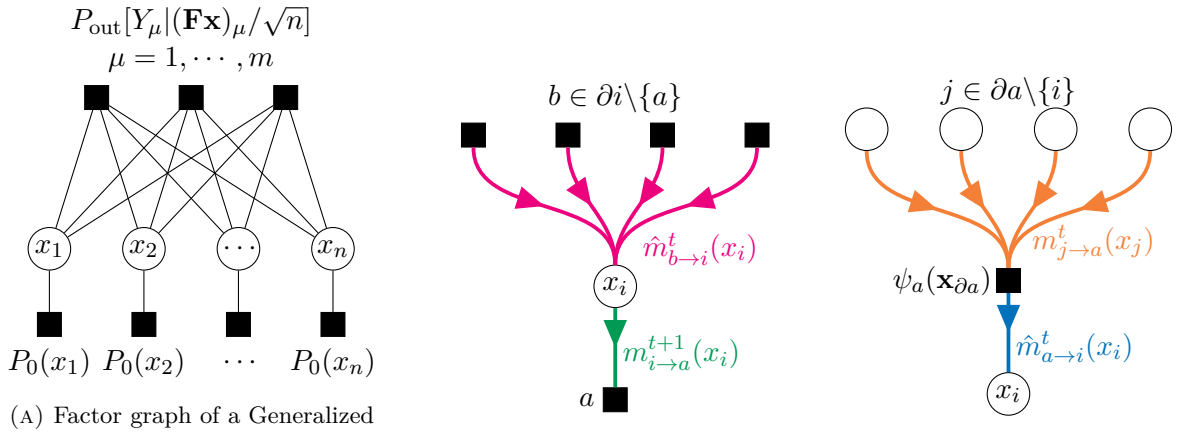
The Generalized Linear Model – For the sake of the presentation, we will focus in Sections 1.4.1 and 1.4.2 on the Generalized Linear Model (Model 1.1), in which the sensing matrix $\mathbf{F}$ is generated i.i.d. from the Gaussian distribution $\mathcal{N}_\beta(0, 1)$. For simplicity we will also consider the real case $\beta = 1$, while the generalization to the complex case is straightforward, see e.g. [Sch16]. The derivation of BP and AMP in this context is quite classic, and the reader can also find them e.g. in [KMS$^+$12, Gab20].

1.4.1 Belief propagation (BP)

Assume we have access to a high-dimensional probability distribution $P(\mathbf{x})$ with $\mathbf{x} \in \mathcal{X}^n$. We would like to compute some important quantities related to this distribution (e.g. the free entropy, as we did in Section 1.3), or to be able to sample in a reasonable computational time. A naive approach of exhausting configurations would require an exponential number of operations $|\mathcal{X}|^n$, which makes it intractable in practice. *Belief propagation* (BP) was introduced precisely to answer these questions in a tractable time – i.e. polynomial in the dimension n. Before diving into the details, let us mention that Chapter D of [MM09] is a very complete introduction to the belief propagation equations that the reader interested in these topics should read.

Factor graph representation

A particularly useful tool to represent high-dimensional probability distributions is *factor graphs*: they are undirected bipartite graphs with two types of nodes, *variable* and *factor* nodes. To fix the ideas, let us consider such a graph G with a set V of variable nodes, F of factor nodes, and let E be the set of edges. Each variable node $i \in V$ represents one of the variables to infer, while each factor node $a \in F$ represents one of the interactions present in the probability distribution.



(A) Factor graph of a Generalized Linear Model, cf. eq. (1.6).

(B) Update of a message $m_{i\to a}$.

(C) Update of a message $\hat{m}_{a\to i}$.

FIGURE 1.7: Representation of a factor graph (Fig. 1.7a) and of the updates of the generic BP equations (1.38) (Figs. 1.7b and 1.7c).

As the graph is bipartite, edges only connect variable and factor nodes, so that the neighbors $\partial a \subseteq V$ of the factor node a are all the variable nodes arising in the interaction term represented by a. All in all, if we denote $\psi_a(\mathbf{x}_{\partial a})$ the interaction represented by a factor node a, a generic factor graph represents the probability distribution

$$P(\mathbf{x}) = \frac{1}{\mathcal{Z}_n} \prod_{a \in F} \psi_a(\mathbf{x}_{\partial a}). \tag{1.37}$$

Notation – Throughout this description we will generically use the notations $\mathbf{x}_A$ for $A \subseteq V$ to designate the subset of variables $\{x_i\}_{i \in A}$.

Conversely, when given an arbitrary probability distribution, one can write it in the form of eq. (1.37), which allows then to draw the corresponding factor graph. As an example, the factor graph representation of the posterior distribution of the GLM (eq. (1.6)) is given in Fig. 1.7a.

Belief propagation equations

Let us consider G a tree factor graph (i.e. with no loops of any size). The Belief Propagation (BP) algorithm is an inference procedure that aims as computing marginals of the underlying distribution $P(\mathbf{x})$, cf. eq (1.37). As we saw, these marginals characterize the MMSE estimator in Bayes-optimal inference problems, see Section 1.1. The BP algorithm uses a set of auxiliary distributions (called *messages*), one for each edge of the graph, that are propagated using precise update rules. Roughly speaking, a message (e.g. $\hat{m}_{a\to i}(x_i)$, as depicted in Fig. 1.7) communicates the *belief* (hence the name of the procedure) that a variable node i takes value x_i, based on all the nodes already visited along the tree. More precisely, the belief propagation update rule is the following[8]:

$$\begin{cases} \hat{m}^t_{a\to i}(x_i) &= \dfrac{1}{\mathcal{Z}_{a\to i}} \displaystyle\sum_{\mathbf{x}_{\partial a \backslash i}} \psi_a(\mathbf{x}_{\partial a}) \prod_{j \in \partial a \backslash i} m^t_{j\to a}(x_j), \\ m^{t+1}_{i\to a}(x_i) &= \dfrac{1}{\mathcal{Z}_{i\to a}} \displaystyle\prod_{b \in \partial j \backslash a} \hat{m}^t_{b\to i}(x_i). \end{cases} \tag{1.38}$$

[8] Eq. (1.38) is sometimes called the *sum-product* update rule, in contrast with the (equivalent) *max-sum* formalism sometimes used, which corresponds in essence to a zero-temperature limit of the BP equations, see [MM09].

Algorithm 1: Sampling procedure using BP iterations [MM09].

Result: A sample $\mathbf{x} \sim P(\mathbf{x})$.
Input: The factor graph G representing $P(\mathbf{x})$, with variable node set V;
Initialize $U = \emptyset$;
for $t \in \{1, \cdots, n\}$ **do**
  Run BP iterations (1.38) until convergence;
  Pick $i \in V \backslash U$;
  Compute the marginal $m_i(x_i)$ by eq. (1.40);
  Draw x_i^* according to $m_i(x_i)$;
  Add a factor $\mathbb{1}\{x_i = x_i^*\}$ in the factor graph ;
  $U \leftarrow U \cup \{i\}$
end
Return $\mathbf{x} \equiv (x_1^*, \cdots, x_n^*)$;

The BP updates of eq. (1.38) are represented in Fig. 1.7b and 1.7c. Note that the sum over configurations naturally becomes an integral in the continuous setting. On tree factor graphs, the BP equations provably describe the probability distribution $P(\mathbf{x})$:

Theorem 1.2 (*Exactness of BP on trees [Pea82, MM09]*)

If the factor graph G is a tree with maximal distance t_* between two variable nodes, then the BP iterations (1.38) provably converge to a stationary point in at most t_* iterations, whatever the initialization. This stationary point is unique, and is an exact representation of the probability distribution $P(\mathbf{x})$ (e.g. for sampling, marginalization and free entropy computation, see below).

The idea behind Theorem 1.2 is that in the updates of eq. (1.38), all the incoming messages are independent from each other (i.e. all the nodes on the top of Figs. 1.7b,c are not connected to each other). In particular, if we remove the node a in Fig. 1.7c, then all the $\{x_j\}_{j\in\partial a}$ become independent under the new factor graph.

Application to the GLM – Even though the factor graph of Fig. 1.7a is clearly not a tree, let us forget this for a moment and apply the BP update (1.38) to the GLM. It yields the following iterative equations:

$$\begin{cases} \hat{m}^t_{\mu\to i}(x_i) &= \dfrac{1}{\mathcal{Z}_{\mu\to i}} \displaystyle\int \Big\{ \prod_{j(\neq i)} \mathrm{d}x_j \, m^t_{j\to\mu}(x_j) \Big\} P_{\text{out}}\Big(y_\mu \Big| \frac{1}{\sqrt{n}} \sum_{k=1}^n F_{\mu k} x_k \Big), \\ m^{t+1}_{i\to\mu}(x_i) &= \dfrac{1}{\mathcal{Z}_{i\to\mu}} P_0(x_i) \displaystyle\prod_{\nu(\neq\mu)} \hat{m}^t_{\nu\to i}(x_i). \end{cases} \tag{1.39}$$

Three applications of the BP algorithm

The belief propagation updates of eq. (1.38) define an iterative algorithm able to solve three important tasks for tree graphical models.

(*i*) **Marginalization** – At their converging point, the BP iterations allow to compute efficiently the marginal distributions of each variable x_i as:

$$m_i(x_i) = \frac{1}{\mathcal{Z}_i} \prod_{a\in\partial i} \hat{m}_{a\to i}(x_i). \tag{1.40}$$

(*ii*) **Sampling** – The BP algorithm also allows to sample from the probability distribution of eq. (1.37), leveraging the computation of the marginals for a fixed set of variables $\mathbf{x}_U$ with $U \subseteq V$. This is summarized in Algorithm 1.

(*iii*) **The Bethe free entropy** – Finally, the BP messages can also be used to compute the free entropy of the model (i.e. the average of the log-normalization factor in eq. (1.37)), as a function of the messages. This function is called the *Bethe free entropy* and is expressed as (see [MM09] for the derivation):

$$\Phi_{\text{Bethe}} \equiv \sum_{i\in V} \ln\left\{\sum_{x_i}\prod_{a\in\partial i}\hat{m}_{a\to i}(x_i)\right\} + \sum_{a\in F}\ln\left\{\sum_{\mathbf{x}_{\partial a}}\psi_a(\mathbf{x}_{\partial a})\prod_{i\in\partial a} m_{i\to a}(x_i)\right\}$$
$$- \sum_{(i,\mu)\in E}\ln\left\{\sum_{x_i} m_{i\to a}(x_i)\hat{m}_{a\to i}(x_i)\right\}. \tag{1.41}$$

Loopy Belief Propagation

As we saw, the BP iterations provably provide access to many important features of the distribution $P(\mathbf{x})$, in polynomial time. However, *a priori* this only stands for tree factor graphs, which limits a lot the range of applicability of this algorithm. As we saw, this limitation arises as the incoming messages in the updates (cf. Fig. 1.7) must be independent. All the rigorous mathematical analysis of BP falls down when the factor graph contains loops, e.g. in Fig. 1.7a representing the GLM.

However, one can still iterate eqs. (1.38), despite the lack of theoretical guarantees. This strategy is generically known as *loopy belief propagation*, and a general theory of it is still missing. Nevertheless, various studies demonstrated that loopy BP could provide very good approximations to the marginals and the free entropy in a variety of loopy factor graphs. For instance, if all loops in the graph are long, the graph will effectively be locally tree-like, and the BP iteration then provide accurate approximations. Loopy BP has also been used in infinite-range inference models [ZK16], and we will focus on this kind of models in the following. Moreover, the BP fixed points can be shown to be related to the stationary points of the Bethe free entropy of eq. (1.41) for any type of factor graph [MM09].

In particular, applying BP to the GLM, the updates of eq. (1.39) are an attractive tool to e.g. sample from the posterior distribution of eq. (1.6), or to compute the asymptotic free entropy via eq. (1.41). As we will see, in the large-n limit these updates will turn out to yield the optimal performance achievable in polynomial time in GLMs, even though the graph of Fig. 1.7a is extremely far from a tree!

1.4.2 Approximate Message Passing (AMP): derivation and consequences

Although easy to write, the BP equations of eq. (1.38) are still computationally very heavy, since one needs to compute the messages for any value of the variables, and that they are $\Theta(n^2)$ messages in long-range models such as the GLM. In [DMM09], the authors proposed a novel algorithm, that they named *Approximate Message Passing* (AMP), for compressed sensing with Gaussian matrices. This algorithm is derived from BP, and it can be seen as a way to make the BP equations tractable, using two features:

- The high-dimensional limit $n \to \infty$, which we will allow us to use "central limit theorem"-type results.

- The projection of the BP messages on a parametrized family of distributions (here Gaussians), reducing the computation of each message to the one of a few scalar quantities. This projection is justified using the high-dimensional limit mentioned above.

This algorithm was later generalized to the GLM of Model 1.1 with Gaussian matrices and arbitrary outputs [Ran11], and later to more generic ensembles of matrices [SRF16, RSF17]. We follow here a very classical route to derive the AMP algorithm from the BP equations, see also the presentations of [ZK16, Gab20][9].

Relaxed belief propagation

The first step of the procedure is to write what is called *relaxed belief propagation* equations, which were derived in [Ran10]. Let us focus on the GLM, that is eq. (1.39). We start by the first equation. Under the statistics of $\prod_{j(\neq i)} \mathrm{d}x_j m_{j\to\mu}(x_j)$, and conditioned on the value of x_i, the variable $\sum_{k(\neq i)} F_{\mu k} x_k$ is approximately Gaussian in the large n limit, with mean and variance given by:

$$\omega^t_{\mu\to i} \equiv \frac{1}{\sqrt{n}} \sum_{j(\neq i)} F_{\mu j} a^{t-1}_{j\to\mu} \quad \text{and} \quad V^t_{\mu\to i} \equiv \frac{1}{n} \sum_{j(\neq i)} F^2_{\mu j} v^{t-1}_{j\to\mu}, \tag{1.42}$$

in which $a_{j\to\mu}, v_{j\to\mu}$ are the mean and variance of x_j under the message $m_{j\to\mu}$. The first equation of eq. (1.39) can then be simplified into

$$\hat{m}^t_{\mu\to i}(x_i) \propto \int \mathrm{d}z_\mu P_{\text{out}}(y_\mu|z_\mu) \exp\Big\{ - \frac{(z_\mu - \omega^t_{\mu\to i} - F_{\mu i} x_i/\sqrt{n})^2}{2V^t_{\mu\to i}} \Big\}. \tag{1.43}$$

Expanding the exponential in eq. (1.43) as $n \to \infty$, we reach that $\hat{m}^t_{\mu\to i}(x_i)$ is approximately Gaussian in this limit, with variance $1/A^t_{\mu\to i}$ and mean $B^t_{\mu\to i}/A^t_{\mu\to i}$ given by:

$$B^t_{\mu\to i} = \frac{1}{\sqrt{n}} F_{\mu i} g_{\text{out}}(y_\mu, \omega^t_{\mu\to i}, V^t_{\mu i}) \quad \text{and} \quad A^t_{\mu\to i} = -\frac{1}{n} F^2_{\mu i} \partial_\omega g_{\text{out}}(y_\mu, \omega^t_{\mu\to i}, V^t_{\mu i}), \tag{1.44}$$

in which we defined:

$$g_{\text{out}}(y, \omega, V) \equiv \frac{1}{V} \frac{\int \mathrm{d}z\, (z-\omega)\, e^{-\frac{(z-\omega)^2}{2V}} P_{\text{out}}(y|z)}{\int \mathrm{d}z\, e^{-\frac{(z-\omega)^2}{2V}} P_{\text{out}}(y|z)}. \tag{1.45}$$

Finally, we focus on the second equation of eq. (1.39). Using the approximate Gaussianity of the messages yields that this equation reduces to:

$$a^t_{i\to\mu} = f_x\Big(\sum_{\nu(\neq\mu)} A^t_{\nu\to i}, \sum_{\nu(\neq\mu)} B^t_{\nu\to i} \Big) \quad \text{and} \quad v^t_{i\to\mu} = \partial_B f_x\Big(\sum_{\nu(\neq\mu)} A^t_{\nu\to i}, \sum_{\nu(\neq\mu)} B^t_{\nu\to i} \Big), \tag{1.46}$$

with the auxiliary function:

$$f_x(B, A) \equiv \frac{\int P_0(\mathrm{d}x)\, x\, e^{-\frac{A}{2}x^2 + Bx}}{\int P_0(\mathrm{d}x)\, e^{-\frac{A}{2}x^2 + Bx}}. \tag{1.47}$$

Eqs. (1.42),(1.44),(1.46) define the relaxed BP updates.

[9] This algorithm is sometimes called *Bayes-AMP* to distinguish it from a broader class of methods also called AMP algorithms. Bayes-AMP has been shown to be optimal in this category of algorithms [CMW20].

Algorithm 2: GAMP for the Generalized Linear Model [Ran11].

Result: The estimator $\hat{\mathbf{x}}$

Input: Observations $\mathbf{Y} \in \mathbb{R}^m$ and sensing matrix $\mathbf{F} \in \mathbb{R}^{m \times n}$;

Initialize $\hat{\mathbf{x}}, \hat{\mathbf{c}}, \boldsymbol{\omega}, \mathbf{V}$ randomly;

while *not converging* **do**

- *Estimation of the mean and variance of* $\boldsymbol{z} \equiv \boldsymbol{F}\hat{\boldsymbol{x}}/\sqrt{n}$;

$$\omega_\mu^t = \frac{1}{\sqrt{n}} \sum_{\mu,i} F_{\mu i} \hat{x}_i^t - \frac{1}{n} g_{\text{out}}(y_\mu, \omega_\mu^{t-1}, V_\mu^{t-1}) \sum_i F_{\mu i}^2 c_i^t \quad \text{and} \quad V_\mu^t = \frac{1}{n} \sum_{\mu,i} F_{\mu i}^2 c_i^t \ ;$$

- *Mean and variance of* $\boldsymbol{x}$ *estimated from the channel observations*;

$$A_i^t = -\frac{1}{n} \sum_\mu F_{\mu i}^2 \, \partial_\omega g_{\text{out}}(y_\mu, \omega_\mu^t, V_\mu^t) \quad \text{and} \quad B_i^t = A_i^t \hat{x}_i^t + \frac{1}{\sqrt{n}} \sum_\mu F_{\mu i} \, g_{\text{out}}(y_\mu, \omega_\mu^t, V_\mu^t) \ ;$$

- *Update of the estimated marginals with the prior information*;

$$\hat{x}_i^{t+1} = f_x(B_i^t, A_i^t) \qquad \text{and} \qquad \hat{c}_i^{t+1} = \partial_B f_x(B_i^t, A_i^t);$$

$t = t + 1$;

end

From relaxed BP to Approximate Message-Passing

The relaxed BP equations can be simplified further in the $n \to \infty$ limit. This is detailed in [ZK16] and essentially reduces to show that all eqs. (1.42),(1.44),(1.46) can be written as a function of the averages over the "output" nodes (e.g. $\omega_\mu \equiv n^{-1} \sum_i \omega_{\mu \to i}$ and $A_i \equiv m^{-1} \sum_\mu A_{i \to \mu}$). One must however be careful when replacing the messages with these averages, to take properly into account all terms of leading order in n. Carrying out this procedure yields Algorithm 2.

Note that there are corrections in Algorithm 2 with what a "naive" average of the r-BP equations (1.42),(1.44),(1.46) would give. These corrections are similar to the Onsager reaction terms arising the TAP approach in Section 1.3.2, and we will see actually that they are completely equivalent to them. Algorithm 2 was first described in this form by Sundeep Rangan in [Ran11], and is a generalization of the original AMP algorithm written by Donoho, Maleki and Montanari for compressed sensing in [DMM09].

Connection with the TAP equations – Let us examine the stationary limit of Algorithm 2. It yields that any fixed point of GAMP must satisfy:

$$\begin{cases} \omega_\mu &= \dfrac{1}{\sqrt{n}} \displaystyle\sum_{\mu,i} F_{\mu i} \hat{x}_i - \frac{1}{n} g_{\text{out}}(y_\mu, \omega_\mu, V_\mu) \sum_i F_{\mu i}^2 c_i \quad \text{and} \quad V_\mu = \frac{1}{n} \sum_{\mu,i} F_{\mu i}^2 c_i, \\ A_i &= -\dfrac{1}{n} \displaystyle\sum_\mu F_{\mu i}^2 \, \partial_\omega g_{\text{out}}(y_\mu, \omega_\mu, V_\mu) \quad \text{and} \quad B_i = A_i \hat{x}_i + \frac{1}{\sqrt{n}} \sum_\mu F_{\mu i} \, g_{\text{out}}(y_\mu, \omega_\mu, V_\mu), \\ \hat{x}_i &= f_x(B_i, A_i) \qquad \text{and} \qquad \hat{c}_i = \partial_B f_x(B_i, A_i). \end{cases} \tag{1.48}$$

In fact, eq. (1.48) are exactly the TAP equations for the GLM, that we introduced in Section 1.3.2 for the SK model. Generalizing the TAP picture from the SK model to the GLM, and subsequently obtaining eq. (1.48), is not very involved and we will detail it in a much more general context in Chapter 2, so that we leave this computation aside.

On a historical note, while the TAP equations have been used by physicists since the 1970s [TAP77], the AMP algorithm (because it is derived from BP) provides an explicit (and non-trivial!) iteration scheme of these equations. This is of the utmost importance: the iteration of the TAP equations in disordered systems is a long-lasting challenge of theoretical physics (see e.g. [Bol14, OCW16] for discussions on the SK model and on more generic Ising models), and

has led to a significant literature [TAP77, Méz89, OW96, BBBZ07, Kab08a][10]. This issue was solved for a large class of inference models by AMP algorithms. The Bethe free entropy (1.41), applied at the fixed point of the AMP algorithm, can also be shown to be equal to the TAP free entropy introduced in Section 1.3.2, and we will detail this correspondence further in Chapter 2.

State evolution and connection to the replica method

An extremely important characteristic of AMP algorithms is that they can be analyzed in the high-dimensional limit. More precisely, we can compute their asymptotic performance via *State Evolution* (SE) equations (as named by [DMM09] for compressed sensing). The most natural way to derive them is to start from the relaxed-BP equations (1.42),(1.44),(1.46). Using the assumption of independence of incoming messages in the BP approach, one can use the central limit theorem to reduce the equations to the evolution of the mean and variances of the sums of many random variables. Tracking the iterations of these means and variances yields the state evolution equations, and a detailed derivation following these lines can be found in [KMS+12]. In the context of the GLM with arbitrary output channel, the SE equations were proven in [JM13], so that we present them here as a theorem. In order to precisely state them, we first describe the results of the replica theory (cf. Section 1.3.1) for the GLM with Gaussian matrices.

Replica formula for the free entropy – As we mentioned in Section 1.3.1, the replica method, initially developed for disordered systems, has proven to be a very powerful tool as well for the analysis of inference models. In particular, the replica-symmetric formula for Bayes-optimal generalized linear models with Gaussian sensing matrices builds on early physics analysis of the perceptron [GD89, Gyö90, STS90], and was derived in [KMS+12]. The formula was eventually proven in [BKM+19], confirming the physics conjecture, so that we refer to this work for the following theorem.

Theorem 1.3 (*RS formula for Gaussian GLMs, informal [BKM+19]*)

Assume that:

- The signal $\mathbf{X}^* \in \mathbb{R}^n$ is generated from an i.i.d. prior distribution P_0, with zero mean and variance $\mathbb{E}_{P_0}[X^2] = \rho > 0$.
- The matrix $\mathbf{F}$ has i.i.d. elements from $\mathcal{N}(0,1)$.
- We are in the Bayes-optimal setting, i.e. the channel and prior used in the posterior distribution of eq. (1.6) are the ones used to generate the data Y_μ and the signal $\mathbf{X}^*$.

Then the free entropy of the posterior distribution of eq. (1.6) converges to:

$$\Phi \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\ln \mathcal{Z}_n(\mathbf{Y},\mathbf{F}) = \sup_{q\in[0,\rho]} \inf_{\hat{q}\geq 0} f_{\rm RS}(q,\hat{q}) = \sup_{q,\hat{q}} f_{\rm RS}(q,\hat{q}), \tag{1.49}$$

with the *replica-symmetric* potential defined as[11]:

$$f_{\rm RS}(q,\hat{q}) \equiv \psi_0(\hat{q}) + \alpha\Psi_{\rm out}(q) - \frac{\hat{q}q}{2}, \tag{1.50}$$

$$\begin{cases} \psi_0(\hat{q}) & \equiv \displaystyle\int \mathcal{D}Z P_0(\mathrm{d}X^*) \exp\Big\{-\frac{\hat{q}}{2}(X^*)^2 + \sqrt{\hat{q}}X^*Z\Big\} \ln \int P_0(\mathrm{d}x)\exp\Big\{-\frac{\hat{q}}{2}x^2 + \sqrt{\hat{q}}xZ\Big\}, \\ \Psi_{\rm out}(q) & \equiv \displaystyle\int \mathrm{d}Y\mathcal{D}V\mathcal{D}W^* P_{\rm out}(Y|\sqrt{q}V + \sqrt{\rho-q}W^*) \ln \int \mathcal{D}w P_{\rm out}(Y|\sqrt{q}V + \sqrt{\rho-q}w). \end{cases}$$

[10] Note that an iteration scheme very close to the AMP algorithm, and derived as well from the BP equations, was first proposed in [KU04].

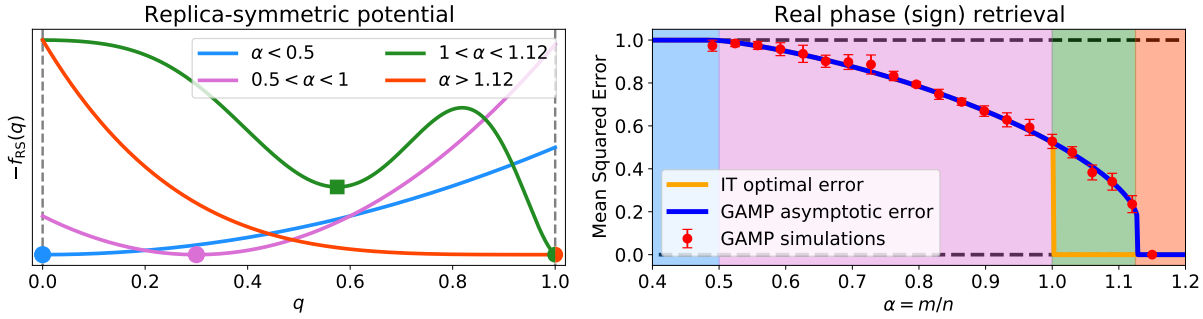


FIGURE 1.8: (Left) Replica-symmetric potential $f_{\rm RS}(q)$ for $\rho = 1$ and different values of the sampling ratio α. We plot the global minimum as a circle and the local minimum closest to $q = 0$ (if different) as a square. The replica potential picture is merely a representation and not an accurate picture. (Right) Mean squared error $\rho - q$ achieved in the *sign retrieval* problem with Gaussian matrices. We compare the global maximum of $f_{\rm RS}$ (information-theoretic MMSE) to the local maximum closest to $q = 0$ (GAMP asymptotic error). We confirm our predictions with finite-size simulations of GAMP. The qualitative behavior of the replica-symmetric potential in each of the four colored areas is given by the curve of the left figure with the corresponding color.

State evolution of AMP – We can now state the SE equations for the GAMP algorithm.

Theorem 1.4 (*Bayes-optimal State Evolution of GAMP, informal [BM11, JM13]*)

Let $\hat{\mathbf{x}}^t$ be the estimator of GAMP (Algorithm 2), and $\mathbf{X}^*$ the solution to infer. We define two asymptotic quantities known as *overlaps*

$$q^t_{\rm AMP} \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (\hat{x}^t_i)^2 \quad \text{and} \quad m^t_{\rm AMP} \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \hat{x}^t_i X^*_i. \tag{1.51}$$

Then these two quantities are well-defined, and moreover by the Nishimori identity (Proposition 1.1) one has $m^t_{\rm AMP} = q^t_{\rm AMP}$ along the trajectory of GAMP. We introduce another variable $\hat{q}_{\rm AMP} \equiv 2\alpha\Psi_{\rm out}(q_{\rm AMP})$. Then at all iteration times $t \geq 0$, $(q_{\rm AMP}, \hat{q}_{\rm AMP})$ is a solution of the *State Evolution (SE)* equations:

$$\hat{q}^t_{\rm AMP} = 2\alpha\Psi_{\rm out}(q^t_{\rm AMP}) \quad \text{and} \quad q^{t+1}_{\rm AMP} = 2\psi_0(\hat{q}^t_{\rm AMP}). \tag{1.52}$$

Moreover, for a random initialization of GAMP we have $q^{t=0}_{\rm AMP} = \hat{q}^{t=0}_{\rm AMP} = 0$.

Optimality of AMP – Theorem 1.4 clarifies the link between AMP and the replica method. It shows precisely that, in the Bayes-optimal setting, the replica potential $f_{\rm RS}$ describes both the information-theoretic optimal performance (via its global maximum, cf. Theorem 1.3) and the asymptotic performance of AMP, via the SE equations (1.52). Note that these equations correspond to a local optimization algorithm of $f_{\rm RS}$ starting from the point $q = \hat{q} = 0$. Basically, combining the replica method and message-passing algorithms allowed us to reduce the study of the fundamental limits of a high-dimensional inference problem to the investigation of the landscape of a simple scalar potential, as illustrated in Fig. 1.8 (see Chapter 6 for more details on the specific problem of phase/sign retrieval illustrated in this figure). Importantly, the link between the state evolution of AMP and the replica-symmetric potential is actually very general, and has several crucial consequences:

- When $f_{\rm RS}$ has a unique maximum in a Bayes-optimal inference model, AMP (here GAMP) achieves the information-theoretic optimal error. However when it is not the case (e.g. the

[11] Recall that $\mathcal{D}$ is the generic notation for the standard Gaussian measure $\mathcal{N}(0, 1)$.

green region in Fig. 1.8-right), Theorem 1.4 leads to the conjecture that AMP should achieve the optimal performance (i.e. the optimal overlap, or MSE) over all "local" polynomial-time algorithms. Indeed, any such algorithm should start at the un-informative point $q = 0$, and would not be able to cross the "barrier" in the replica-symmetric potential in polynomial time. Of course this argument is very far from rigorous and such a wild statement should not be taken as is. There indeed exists problems for which this statement is not correct, notably the famous XOR-SAT problem for which all local search algorithms (including AMP) fail, while the problem can be solved in polynomial time by Gaussian elimination using a well-designed mapping [RTWZ01, BHL+02, JMS04]. In an effort to put the conjectured optimality of AMP on more precise grounds, the authors of [CMW20] have shown that AMP is optimal in a very large class of inference problems among all *general first order methods*, which includes e.g. all gradient-based optimization methods. This result covers in particular the GLM, and it proves that Fig. 1.8 indeed describes the fundamental statistical and algorithmic performances in real Gaussian phase retrieval *among general first order methods*.

- To summarize, the replica formula (here Theorem 1.3) gives access to the information-theoretic optimal error, while the class of AMP algorithms provides the (conjectured) algorithmic optimal error, and this error is also characterized by the replica formula (see here Theorem 1.4). This allows to study *computational-to-statistical gaps* (also called hard phases) merely by analyzing the replica-symmetric potential, a simple function of a few scalars! Leveraging this statistical physics toolbox to study computational gaps in inference is an important line of research, see e.g. [KMS+12, DJM13, DM15, ZK16, BPW18, BKM+19] (without any aim of exhaustivity), and a significant fraction of Part II of the present thesis will be devoted to the investigation of such gaps.

Final remarks on AMP algorithms

AMP for spin glass optimization – AMP algorithms have recently been adapted as optimization algorithms in a variety of spin glass models with full replica symmetry breaking (see Section 1.3.1). In this context they were called *Incremental AMP*, and they have solved long-lasting open problems on the optimization of the SK model and a wide class of FRSB spin glass models satisfying a "no-overlap-gap" property [AMS20, AM20, Sub21, Mon21, Sel21].

The cavity method – In our presentation of methods which originated in statistical physics, we did not detail a very important technique known as the *cavity method*. It was introduced by Mézard and coauthors in [MPV86] as a way to obtain the "replica solution without replicas" for the SK model. The basis of the cavity method for generic graphical models described by eq. (1.37) is to study the original model along with a modified version in which one removed a single variable node: as these two systems must be equivalent in the thermodynamic limit this yields self-consistent equations on the order parameters of the system. For the case of the GLM, these self-consistent equations are exactly the state evolution of Theorem 1.4! The cavity method has actually been shown generically to be equivalent to the replica computations[12], and can be generalized to an arbitrary number of replica symmetry breaking levels. As we will not discuss the cavity computations in this thesis, the reader interested in learning more should refer to [MM09, Méz15].

The mismatched case – The BP equations (1.39) and the GAMP algorithm 2 do not use the knowledge of the true prior and channels, so that they are easily generalized to the non-Bayes-optimal (or "mismatched") case. The state evolution (Theorem 1.4) can also be generalized to this more general case, and is again equivalent to the replica-symmetric computation. However,

[12] However cavity computations are often easier to prove mathematically than their replica counterpart, as they rely on less heuristic methods.

out of Bayes-optimality, the replica-symmetric computation generally does not yield the correct asymptotic free entropy of the system, so that we have no guarantee of any kind of optimality of AMP algorithms. However as we will illustrate in Chapter 6, these algorithms can still achieve very good performance in the mismatched case. As we will most often need the Bayes-optimal version of the AMP algorithm in this dissertation we do not present here the mismatched state evolution, which the reader can find e.g. in [KMS+12] or [Gab20].

1.4.3 Three approximations for non-Gaussian inference problems

In this section, inspired by [MFC+19], we introduce three different approximation schemes for disordered systems that can also be applied to many inference problems, when the underlying interactions are more structured than simply Gaussian. Let us briefly recap their history:

- The *adaptive TAP* (adaTAP) scheme was developed and presented in 2001 in [OW01b, OW01a, OS01], for systems close to the SK model.
- The same year, Thomas Minka's *Expectation Propagation* (EP) approach was presented [Min01]. Opper and Winther used an alternative view of local-consistency approximations of the EP–type which they call *Expectation Consistent* (EC) approximations in [OW05a, OW05b], effectively re-deriving their adaTAP scheme from this new point of view.
- The *Vector Approximate Message-Passing* (VAMP) approach is more recent [SRF16, RSF17], and is again another EP approach, for Generalized Linear Models (GLMs). Compared with other EP-like approaches it has the advantage that it leads to both a practical converging algorithm, and a rigorous treatment of its time evolution.

The connection between these different approaches was hinted several times for SK-like problems, see e.g. [OCW16, ÇO19], and we establish it clearly in the remaining of Section 1.4.3, along with a detailed presentation of these three approximations.

Expectation Consistency approximation

Consider a model in which the probability distribution of a vector $\mathbf{x} \in \mathbb{R}^n$ is of the generic form:

$$P(\mathbf{x}) = \frac{1}{\mathcal{Z}_n} P_0(\mathbf{x}) P_J(\mathbf{x}). \tag{1.53}$$

As we saw above, such distributions typically appear in Bayesian approaches to inference problems, and as emphasized in Section 1.1, our goal will be to compute $\ln \mathcal{Z}_n$ for large values of n. We will use the Bayesian language and denote P_0 as a *prior* distribution on $\mathbf{x}$, which will be typically factorized (i.e. all the components of $\mathbf{x}$ are independent under P_0). The distribution P_J is responsible for the interactions between the $\{x_i\}$: we will often be interested in *pairwise interactions*, i.e. when $\ln P_J$ is a quadratic form in the $\{x_i\}$ variables. An example of such a model is the infinite-range Ising model at inverse temperature $\eta \geq 0$, with external field $\mathbf{h}$ (i.e. a SK model with external field and non-Gaussian couplings):

$$\begin{cases} P_0(\mathbf{x}) &= \displaystyle\prod_{i=1}^{n} \Big\{ \frac{1}{2\cosh(\eta h_i)} \Big(\delta(x_i - 1) e^{-\eta h_i} + \delta(x_i + 1) e^{\eta h_i} \Big) \Big\}, \\ P_J(\mathbf{x}) &= \exp\Big\{ \frac{\eta}{2} \sum_{i,j} J_{ij} x_i x_j \Big\}. \end{cases} \tag{1.54}$$

Each of the two distributions P_0 and P_J allows for tractable computations of physical quantities (e.g. averages), but the difficulty arises when considering their product. The idea behind EC is to simultaneously approximate P_0 and P_J by a tractable family of distributions. For the sake

of the presentation we will consider Gaussian probability distributions, although this can be generalized to different families, as detailed in [OW05a]. We define the first approximation as:

$$\mu_0(\mathbf{x}) \equiv \frac{1}{\mathcal{Z}_0(\mathbf{\Gamma}_0, \boldsymbol{\lambda}_0)} P_0(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}},$$

with $\mathbf{\Gamma}_0 \in \mathcal{S}_n^+(\mathbb{R})$ and $\boldsymbol{\lambda}_0 \in \mathbb{R}^n$. We will denote $\langle \cdot \rangle_0$ the averages with respect to μ_0. We can write the trivial identity:

$$\mathcal{Z}_n = \mathcal{Z}_n \times \frac{\mathcal{Z}_0(\mathbf{\Gamma}_0, \boldsymbol{\lambda}_0)}{\mathcal{Z}_0(\mathbf{\Gamma}_0, \boldsymbol{\lambda}_0)} = \mathcal{Z}_0(\mathbf{\Gamma}_0, \boldsymbol{\lambda}_0) \langle P_J(\mathbf{x})\, e^{\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_0 \mathbf{x} - \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \rangle_0.$$

When computing the average $\langle P_J(\mathbf{x})\, e^{\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_0 \mathbf{x} - \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \rangle_0$, we can replace the distribution μ_0 by an approximate Gaussian distribution:

$$\mu_S(\mathbf{x}) \equiv \frac{1}{\mathcal{Z}_S} e^{-\frac{1}{2}\mathbf{x}^\intercal (\mathbf{\Gamma}_J + \mathbf{\Gamma}_0)\mathbf{x} + (\boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_J)^\intercal \mathbf{x}}.$$

This yields the *Expectation-Consistency* (EC) approximation to the free entropy:

$$\begin{aligned} \ln \mathcal{Z}^{\mathrm{EC}}(\mathbf{\Gamma}_0, \mathbf{\Gamma}_J, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_J) = \ln \Big[\int \mathrm{d}\mathbf{x} P_0(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \Big] + \ln \Big[\int \mathrm{d}\mathbf{x} P_J(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}} \Big] \\ - \ln \Big[\int \mathrm{d}\mathbf{x}\, e^{-\frac{1}{2}\mathbf{x}^\intercal (\mathbf{\Gamma}_0 + \mathbf{\Gamma}_J)\mathbf{x} + (\boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_J)^\intercal \mathbf{x}} \Big]. \end{aligned} \tag{1.55}$$

Note that all three parts of this free entropy are tractable. In order to symmetrize the result we can define a third measure μ_J with average $\langle \cdot \rangle_J$:

$$\mu_J(\mathbf{x}) \equiv \frac{1}{\mathcal{Z}_J(\mathbf{\Gamma}_J, \boldsymbol{\lambda}_J)} P_J(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}}.$$

The final free entropy should not depend on the values of the parameters, so we expect that the optimal values for $\mathbf{\Gamma}_0, \mathbf{\Gamma}_J, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_J$ make $\mathcal{Z}^{\mathrm{EC}}$ stationary. This is a strong hypothesis, and the reader can refer to [OW05a] for more details and justifications. These stationarity equations yield the *Expectation Consistency* (EC) conditions, giving their name to the procedure:

$$\begin{cases} \langle x_i \rangle_0 & = \langle x_i \rangle_J = \langle x_i \rangle_S, \\ \langle x_i x_j \rangle_0 & = \langle x_i x_j \rangle_J = \langle x_i x_j \rangle_S. \end{cases} \tag{1.56}$$

Adaptive TAP approximation

The adaptive TAP (adaTAP) approximation [OW01a, OW01b] provides an equivalent way to derive the free entropy of eq. (1.55) for models with pairwise interactions. Let us briefly sketch its derivation and the main arguments behind it. Again, for the sake of the presentation we consider the infinite-range Ising model of eq. (1.54). As we saw in Section 1.3.2, the free entropy $\Phi_{\mathbf{J}} \equiv n^{-1} \ln \mathcal{Z}_n$ at fixed values of the magnetizations $m_i = \langle x_i \rangle$ and $v_{ij} = \langle x_i x_j \rangle$ can be written using Lagrange multipliers $(\boldsymbol{\lambda}, \mathbf{\Gamma})$:

$$\Phi_{\mathbf{J}}(\eta, \mathbf{m}, \mathbf{v}) = \underset{\substack{\boldsymbol{\lambda} \in \mathbb{R}^n \\ \mathbf{\Gamma} \in \mathcal{S}_n^+(\mathbb{R})}}{\mathrm{extr}} \Big[-\boldsymbol{\lambda}^\intercal \mathbf{m} + \sum_{i,j} \frac{\Gamma_{ij}}{2} (v_{ij} + m_i m_j) + \ln \int \mathrm{d}\mathbf{x}\, P_0(\mathbf{x})\, e^{\frac{\eta}{2}\mathbf{x}^\intercal \mathbf{J} \mathbf{x} - \frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma} \mathbf{x} + \boldsymbol{\lambda}^\intercal \mathbf{x}} \Big].$$

The adaTAP approximation consists in writing:

$$\begin{aligned} n\Phi_{\mathbf{J}}(\eta,\mathbf{m},\mathbf{v}) &= \Phi_{\mathbf{J}}(0,\mathbf{m},\mathbf{v}) + \int_0^\eta \mathrm{d}l\, \frac{\partial \Phi_{\mathbf{J}}(l,\mathbf{m},\mathbf{v})}{\partial l}, \\ &\simeq \Phi_{\mathbf{J}}(0,\mathbf{m},\mathbf{v}) + \int_0^\eta \mathrm{d}l\, \frac{\partial \Phi_{\mathbf{J}}^{(\text{Gauss.})}(l,\mathbf{m},\mathbf{v})}{\partial l}, \\ &\simeq \Phi_{\mathbf{J}}(0,\mathbf{m},\mathbf{v}) + \Phi_{\mathbf{J}}^{(\text{Gauss.})}(\eta,\mathbf{m},\mathbf{v}) - \Phi_{\mathbf{J}}^{(\text{Gauss.})}(0,\mathbf{m},\mathbf{v}). \end{aligned} \tag{1.57}$$

In this expression $\Phi_{\mathbf{J}}^{(\text{Gauss.})}(\eta,\mathbf{m},\mathbf{v})$ denotes the free entropy of the same system, but in which the spins have *Gaussian statistics*. The idea behind the adaTAP approximation is that in $\partial_l \Phi_{\mathbf{J}}(l,\mathbf{m},\mathbf{v}) = (2n)^{-1}\sum_{ij} J_{ij}\langle x_i x_j\rangle$ it is reasonable to assume that the underlying variables were Gaussian, as we consider the expectation of a sum of a large number of variables. The assumptions of adaTAP, although reasonable, are *a priori* hard to justify more rigorously and systematically. Note that the free entropy (1.57) of adaTAP is equivalent to the one derived using Expectation Consistency in eq. (1.55). Indeed, using additional Lagrange multipliers we can write the three terms of eq. (1.57) as:

$$\begin{aligned} n\Phi_{\mathbf{J}}^{\text{adaTAP}}(\eta,\mathbf{m},\mathbf{v}) = &\operatorname*{extr}_{\boldsymbol{\lambda}_0,\boldsymbol{\Gamma}_0}\left[\ln\left\{\int \mathrm{d}\mathbf{x} P_0(\mathbf{x})\, e^{-\frac{1}{2}\mathbf{x}^\intercal\boldsymbol{\Gamma}_0\mathbf{x}+\boldsymbol{\lambda}_0^\intercal\mathbf{x}}\right\} - \boldsymbol{\lambda}_0^\intercal\mathbf{m} + \frac{1}{2}\sum_{i,j}(\Gamma_0)_{ij}(v_{ij}+m_i m_j)\right] \\ &+ \operatorname*{extr}_{\boldsymbol{\lambda}_J,\boldsymbol{\Gamma}_J}\left[\ln\left\{\int \mathrm{d}\mathbf{x} P_J(\mathbf{x})\, e^{-\frac{1}{2}\mathbf{x}^\intercal\boldsymbol{\Gamma}_J\mathbf{x}+\boldsymbol{\lambda}_J^\intercal\mathbf{x}}\right\} - \boldsymbol{\lambda}_J^\intercal\mathbf{m} + \frac{1}{2}\sum_{i,j}(\Gamma_J)_{ij}(v_{ij}+m_i m_j)\right] \\ &- \operatorname*{extr}_{\boldsymbol{\lambda}_S,\boldsymbol{\Gamma}_S}\left[\ln\left\{\int \mathrm{d}\mathbf{x}\, e^{-\frac{1}{2}\mathbf{x}^\intercal\boldsymbol{\Gamma}_S\mathbf{x}+\boldsymbol{\lambda}_S^\intercal\mathbf{x}}\right\} - \boldsymbol{\lambda}_S^\intercal\mathbf{m} + \frac{1}{2}\sum_{i,j}(\Gamma_S)_{ij}(v_{ij}+m_i m_j)\right]. \end{aligned}$$

Once written in this form, the extremization over $\mathbf{m}$ and $\mathbf{v}$ of the free entropy implies that $\boldsymbol{\Gamma}_S = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_J$ and $\boldsymbol{\lambda}_S = \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_J$. It is then clear that we found back $\ln \mathcal{Z}^{\text{EC}}$ of eq. (1.55).

Vector Approximate Message Passing approximation

The Vector Approximate Message Passing (VAMP) algorithm [RSF17, SRF16] extends previous message-passing approaches, such as the AMP algorithm that we saw in Section 1.4.2 for Gaussian interactions, to a class of correlated interaction matrices that satisfy a *right-rotation invariance property*, that we will define more precisely in Section 1.5. The algorithm itself can be derived in several ways: here we will use a method based on the belief propagation equations on a "duplicated" factor graph, and their Gaussian projection. As we shall see, the Bethe free entropy of this model is then equivalent to the Expectation-Consistency free entropy. For the sake of the presentation we consider again the infinite-range Ising model of eq. (1.54). The idea behind VAMP is to "duplicate" the model as follows:

$$\Phi_{\mathbf{J}} \equiv \frac{1}{n}\ln\int \mathrm{d}\mathbf{x}\, P_0(\mathbf{x}) P_J(\mathbf{x}) = \frac{1}{n}\ln\int \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, P_0(\mathbf{x}_1) P_J(\mathbf{x}_2)\, \delta(\mathbf{x}_1-\mathbf{x}_2).$$

This partition function can be represented as a duplicated factor graph involving two vector nodes, as shown in Fig. 1.9. One then writes the BP equations for this problem, as we saw in Section 1.4.1, in terms of two "messages" $\mathbf{m}_0(\mathbf{x}_2)$ and $\mathbf{m}_J(\mathbf{x}_1)$. These equations can be obtained by the stationarity conditions of the *Bethe free entropy* defined in eq. (1.41), and which reads here:

$$\Phi_{\text{Bethe}} \equiv \ln\int \mathrm{d}\mathbf{x}\, P_0(\mathbf{x})\, \mathbf{m}_J(\mathbf{x}) + \ln\int \mathrm{d}\mathbf{x}\, P_J(\mathbf{x})\, \mathbf{m}_0(\mathbf{x}) - \ln\int \mathrm{d}\mathbf{x}\, \mathbf{m}_0(\mathbf{x})\, \mathbf{m}_J(\mathbf{x}). \tag{1.58}$$

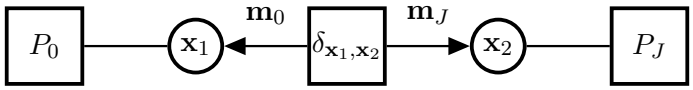


FIGURE 1.9: Duplicated factor graph for the VAMP approximation. Circles represent vector nodes and squares factor nodes. We represent the two messages $\mathbf{m}_0$ and $\mathbf{m}_J$ in terms of which we write the full BP equations.

As the factor graph of Fig. 1.9 is a tree, the BP equations are an exact representation of the original problem, but they are in general intractable. In order to make the computation possible one can make a Gaussian approximation, which is at the core of the VAMP algorithm: the messages $\mathbf{m}_0$ and $\mathbf{m}_J$ are assumed to be Gaussian, and thus are fully characterized by their first two moments:

$$\mathbf{m}_0(\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}}, \qquad \mathbf{m}_J(\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{x}^\intercal \mathbf{\Gamma}_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}}. \tag{1.59}$$

Writing the BP update rule with this assumption yields the VAMP iterations:

$$\begin{cases} \mathbf{\Gamma}_0^t &= \left[\langle \mathbf{x}\mathbf{x}^\intercal \rangle_0^{t-1} - \langle \mathbf{x} \rangle_0^{t-1} \langle \mathbf{x}^\intercal \rangle_0^{t-1}\right] - \mathbf{\Gamma}_J^{t-1}, \\ \boldsymbol{\lambda}_0^t &= (\mathbf{\Gamma}_0^t + \mathbf{\Gamma}_J^{t-1}) \langle \mathbf{x} \rangle_{\mu_0}^{t-1} - \boldsymbol{\lambda}_J^{t-1}, \\ \mathbf{\Gamma}_J^t &= \left[\langle \mathbf{x}\mathbf{x}^\intercal \rangle_J^{t} - \langle \mathbf{x} \rangle_J^{t} \langle \mathbf{x}^\intercal \rangle_J^{t}\right] - \mathbf{\Gamma}_0^t, \\ \boldsymbol{\lambda}_J^t &= (\mathbf{\Gamma}_0^t + \mathbf{\Gamma}_J^t) \langle \mathbf{x} \rangle_{\mu_J}^{t} - \boldsymbol{\lambda}_0^t, \end{cases} \tag{1.60}$$

with the measures $\mu_0(\mathbf{x}) \propto P_0(\mathbf{x})\mathbf{m}_J(\mathbf{x})$ and $\mu_J(\mathbf{x}) \propto P_J(\mathbf{x})\mathbf{m}_0(\mathbf{x})$. Note that plugging the ansatz of eq. (1.59) into eq. (1.58) immediately gives back the EC free entropy of eq. (1.55). Moreover, the stationary limit of the BP equations of eq. (1.60) (i.e. when removing the time indices) is identical to the Expectation-Consistency conditions of eq. (1.56).

Conclusion

We introduced three approximation schemes to compute the free entropy of disordered systems, that have natural extensions in inference problems: Expectation Consistency, adaptive TAP, and Vector Approximate Message Passing. As we detailed, they all rely on the same underlying Gaussian approximations, and therefore yield equivalent expressions for the free entropy. As a final note, an important advantage of the VAMP approach is that it naturally provides an iterative scheme to solve the fixed point equations. These iterations form a very efficient algorithm [RSF17], which is intuitive as they were derived directly from the belief propagation equations.

1.5 Some rudiments of probability and random matrix theory

Theoretical physics and random matrix theory share a long history that dates back to Wigner [Wig55], and that powered progress in various areas ranging from disordered systems [EA75, SK75] to quantum chaos [BGS84], quantum chromodynamics [VW00], or superconductivity [Bah96]. The growing interplay of physics and statistics [ZK16, Gab20, Zde20], which is the subject of this thesis, further strengthened this connection. In this section, the last one of Chapter 1, we review some important techniques and models from random matrix theory, as well as a few probabilistic tools that will be useful for the rest of the thesis. Apart from some specific arguments, it is written in a fashion closer to mathematical standards, as a large

part of its content originated in the mathematics literature. For more details and advanced results in random matrix theory, one can refer for instance to [Meh04] and [AGZ10] for classical references of theoretical physics and mathematics. On a less advanced level, [LNV18] provides a comprehensive introduction to techniques of random matrix theory used in theoretical physics.

1.5.1 Random matrix ensembles and asymptotic spectra

Recall that $\mathcal{M}_1^+(\mathbb{R})$ is the set of probability measures on $\mathbb{R}$.

Spectral distributions – Consider a (random) symmetric matrix $\mathbf{M} \in \mathcal{H}_n(\mathbb{K})$, with (real) eigenvalues $\{\lambda_i\}_{i=1}^n$. The *Empirical Spectral Distribution* (ESD) of $\mathbf{M}$ is defined as:

$$\nu_n \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}. \tag{1.61}$$

For many random matrices considered in this thesis, the (random) probability measure ν_n will converge almost surely and in the weak sense to a deterministic probability measure $\nu \in \mathcal{M}_1^+(\mathbb{R})$ as $n \to \infty$. In this case, we will call ν the *Limiting Spectral Distribution* (LSD) of $\mathbf{M}$.

A few random matrix ensembles

Wigner matrices – In the 1950s, Wigner introduced a class of random matrices to study the nuclei of heavy atoms [Wig55]. This led to what is now known as *Wigner matrices*, which are Hermitian/symmetric random matrices $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ such that:

(*i*) All $\{J_{ij}\}_{i<j}$ are i.i.d. with zero mean and $\mathbb{E}|J_{ij}|^2 = 1$.

(*ii*) All $\{J_{ii}\}_{i=1}^n$ are i.i.d. with zero mean, independent of $\{J_{ij}\}_{i<j}$, and $\mathbb{E}|J_{ii}|^2 = 2$.

Note that if the distribution of the elements is Gaussian, then Wigner matrices are (up to a rescaling) the GOE/GUE matrices of Definition 1.3. The LSD of Wigner matrices is given in Theorem 1.6 and was first investigated in [Wig55]: it can be considered as the beginning of the field of random matrix theory.

(Generalized) Wishart matrices – Wishart matrices, introduced in [Wis28], are a model of *sample covariance matrices*. More precisely, let $\mathbf{x}_1, \cdots, \mathbf{x}_m \in \mathbb{R}^n$ be i.i.d. vectors drawn from a zero-mean distribution $P_{\mathbf{x}} \in \mathcal{M}_1^+(\mathbb{R}^n)$. Their sample covariance matrix is

$$\mathbf{W} \equiv \frac{1}{m}\sum_{\mu=1}^{m} \mathbf{x}_\mu \mathbf{x}_\mu^\intercal \in \mathcal{S}_n.$$

At fixed n, when the number m of data samples gets large, $\mathbf{W}$ approaches the covariance $\mathbb{E}_{P_{\mathbf{x}}}[\mathbf{x}\mathbf{x}^\intercal]$ of the distribution, by the law of large numbers. However the situation is much more complex in a high-dimensional regime, in which both n and m gets large in the same scale, that is $m/n \to \alpha > 0$, a limit which we referred to as the *thermodynamic limit*, cf. Definition 1.2. Let us assume that $P_{\mathbf{x}}$ is a Gaussian distribution with covariance matrix $\mathbf{\Sigma} \in \mathcal{S}_n^{++}(\mathbb{R})$. As $\mathbf{x}_\mu \overset{\mathrm{d}}{=} \sqrt{\mathbf{\Sigma}}\mathbf{z}_\mu$, with $\mathbf{z}_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_n)$, we reach

$$\mathbf{W} \overset{\mathrm{d}}{=} \sqrt{\mathbf{\Sigma}}\left\{\frac{1}{m}\sum_{\mu=1}^{m} \mathbf{z}_\mu \mathbf{z}_\mu^\intercal\right\}\sqrt{\mathbf{\Sigma}}. \tag{1.62}$$

Eq. (1.62) defines the random matrix ensemble of *Wishart matrices*. Under suitable assumptions on $\mathbf{\Sigma}$ their asymptotic spectra was first studied by Marchenko and Pastur in [MP67], see

Theorem 1.7. In Chapter 8 we will consider an extension of Wishart matrices, that we call *generalized sample covariance matrices*.

Rotationally-invariant matrices – Roughly speaking, when we refer to rotational invariance, we mean a complete delocalization of eigenvectors: they are distributed according to the Haar measure on the orthogonal group $\mathcal{U}_\beta(n)$, with no privileged direction. Throughout the manuscript, we will use two assumptions that we will both refer to as rotation invariance.

Model S (*Hermitian/Symmetric rotationally invariant matrix*)

Let $n \geq 1$. The random matrix $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ is said to be **rotationally invariant** if:

(*i*) For any $\mathbf{O} \in \mathcal{U}_\beta(n)$, $\mathbf{J} \overset{\mathrm{d}}{=} \mathbf{OJO}^\dagger$.

(*ii*) The empirical spectral distribution (ESD) of $\mathbf{J}$: $\rho_n \equiv \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(\mathbf{J})}$ converges (almost surely and in the weak sense) as $n \to \infty$ to a probability distribution ρ with compact support, called the Limiting Spectral Distribution (LSD) of $\mathbf{J}$.

(*iii*) The smallest and largest eigenvalue of $\mathbf{J}$ converge (almost surely) to the infimum and supremum of the support of ρ.

Model R (*Rectangular rotationally invariant matrix*)

Let $n \geq 1$, and $m = m(n) \geq 1$ such that $m/n \to \alpha > 0$ as $n \to \infty$.

- The random matrix $\mathbf{L} \in \mathbb{R}^{m \times n}$ is said to be **right-rotationally invariant** if $\mathbf{J} \equiv \mathbf{L}^\dagger\mathbf{L}$ is rotationally invariant, according to Model S.
- In the same way, $\mathbf{L}$ is **left-rotationally invariant** if $\mathbf{J} \equiv \mathbf{L}\mathbf{L}^\dagger$ is a symmetric rotationally invariant matrix.

Note that in both models we added additional hypotheses to the complete delocalization of eigenvectors, more precisely on the asymptotic behavior of the spectral distribution and on the existence of outliers in the spectrum. These hypotheses have been added for convenience as we will use them repeatedly in this manuscript. Note that a strong form of eigenvectors delocalization can also be proven for non-rotationally invariant random matrices, e.g. the Wigner matrices with generic i.i.d. distribution [ESY09].

Examples of rotationally-invariant ensembles –

- The GOE/GUE (cf. Definition 1.3) satisfies Model S, with LSD given by the *Wigner semicircle law* as detailed in Theorem 1.6.
- The Wishart ensemble of eq. (1.62) with $\mathbf{\Sigma} = \mathrm{I}_n$ and ratio $\alpha > 0$. This ensemble satisfies Model S, with LSD given by the *Marchenko-Pastur law* as shown in Theorem 1.7.
- Standard Gaussian i.i.d. matrices $\mathbf{L} \in \mathbb{K}^{m \times n}$ satisfy Model R and are left and right-rotationally invariant.
- Matrices generated via a *potential* $V(x)$, i.e. $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ generated with a probability density proportional to $\exp\{-n \operatorname{Tr}[V(\mathbf{J})]\}$. Given some natural hypotheses on the behavior of the potential $V(x)$, such matrices satisfy Model S.

Useful transforms of probability measures

We now introduce a few transforms of probability measures that are especially useful in the context of random matrix theory. Let $\mathbb{C}_+ = \{z \in \mathbb{C}, \operatorname{Im} z > 0\}$. We let $\nu \in \mathcal{M}_1^+(\mathbb{R})$, and

we denote $\mathrm{supp}(\nu) \subseteq \mathbb{R}$ its support. We define $(t_{\min}, t_{\max})$ as the infimum and supremum of $\mathrm{supp}(\nu)$. For any z such that $z \in \mathbb{C}_+$ or $z \in \mathbb{R}\backslash\mathrm{supp}(\nu)$, we define the *Stieltjes transform* of ν (or *resolvent*) as[13]:

$$\mathcal{S}_\nu(z) \equiv \int \nu(\mathrm{d}t) \frac{1}{t-z}. \tag{1.63}$$

Note that $\mathcal{S}_\nu(z)$ is a one-to-one mapping of $\mathbb{C}_+$ on itself. The Stieltjes transform has proven to be a very useful tool in particular in random matrix theory [AGZ10]. Importantly, the knowledge of the Stieltjes transform of a probability measure above the real line allows to characterize it completely, via the Stieltjes-Perron inversion formula (see Theorem X.6.1 of [DS67]):

Theorem 1.5 (*Stieltjes-Perron inversion formula*)

Let $\nu \in \mathcal{M}_1^+(\mathbb{R})$. Then for all $a < b$, we have

$$\nu((a,b)) = \lim_{\delta \downarrow 0} \lim_{\epsilon \downarrow 0} \frac{1}{2i\pi} \int_{a+\delta}^{b-\delta} [S_\nu(x+i\epsilon) - S_\nu(x-i\epsilon)]\mathrm{d}x.$$

In particular, if ν has a continuous density with respect to the Lebesgue measure then:

$$\forall x \in \mathbb{R}, \quad \frac{\mathrm{d}\nu}{\mathrm{d}x} = \lim_{\epsilon \downarrow 0} \frac{1}{\pi} \mathrm{Im}\, \mathcal{S}_\nu(x+i\epsilon).$$

Let us now assume that $t_{\max} < \infty$, so that the support of ν is bounded from above. On $(t_{\max}, +\infty)$, $\mathcal{S}_\nu$ induces a strictly increasing $\mathcal{C}^\infty$ diffeomorphism $\mathcal{S}_\nu : (t_{\max}, \infty) \hookrightarrow (-\infty, 0)$, and we denote its inverse $\mathcal{S}_\nu^{-1}$. It is easy to see that the same property holds as well on $(-\infty, t_{\min})$. One can then introduce the *R-transform* of ν as:

$$\mathcal{R}_\nu(s) \equiv \mathcal{S}_\nu^{-1}(-s) - \frac{1}{s}. \tag{1.64}$$

$\mathcal{R}_\nu(s)$ is defined for $-s \in \mathcal{S}_\nu[(t_{\min}, t_{\max})^c]$, and one can show that it admits an analytical expansion around $s = 0$, see e.g. [TV04]. We can write this expansion as:

$$\mathcal{R}_\nu(s) = \sum_{k=0}^{\infty} c_{k+1}(\nu)\, s^k. \tag{1.65}$$

The elements $\{c_k(\nu)\}_{k\geq 1}$ are called the *free cumulants* of ν. They are analogous to the usual cumulants since the R-transform is analogous to the cumulant generating function in the context of random matrix theory. This relation was shown in detail in [GM05] and we will detail many results of this work in the following. In particular, one verifies that $c_1(\nu) = \mathbb{E}_\nu[X]$ and $c_2(\nu) = \mathbb{E}_\nu[X^2] - (\mathbb{E}_\nu X)^2$. The free cumulants can be recursively computed from the moments of the measure using the so-called *free cumulant equation*:

$$\forall k \in \mathbb{N}^*, \quad \mathbb{E}_\nu[X^k] = \sum_{m=1}^{k} c_m(\nu) \sum_{\substack{\{k_i\}_{i\in[|1,m|]} \\ \text{s.t } \sum_i k_i = k}} \prod_{i=1}^{m} \mathbb{E}_\nu[X^{k_i - 1}]. \tag{1.66}$$

The free cumulants naturally arise from the theory of *free probability*, which is the study of non-commutative random variables [VDN92]. This field has very strong connections with random matrix theory, and many of the results we will mention can be understood from the free

[13] At different points in this thesis we will use the different notations $g_\nu(z) = \mathcal{S}_\nu(z)$ and $G_\nu(z) = -\mathcal{S}_\nu(z)$.

probability point of view.

Limit spectra of random matrices

As we mentioned, one of the first object of studies of random matrix theory was the limit spectra of different random matrix ensembles. Naturally, the first one to be studied was the Wigner ensemble that we defined above. Its LSD was shown in [Wig55], using a moments method, to be equal to the now-celebrated semicircle distribution.

Theorem 1.6 (*Wigner's semicircle law, [Wig55]*)

Let $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ be a Wigner matrix (see the definition above). Then as $n \to \infty$, the ESD of $\mathbf{J}/\sqrt{n}$ a.s. tends (in the weak sense) to the *semicircle* law $\sigma_{\text{s.c.}}$ with density

$$\sigma_{\text{s.c.}}(x) = \frac{\sqrt{4-x^2}}{2\pi} \mathbb{1}\{|x| \leq 2\}. \tag{1.67}$$

The proof of Theorem 1.6 was performed in [Wig55] using the calculation of all moments of a Wigner matrix. This proof was then much simplified by use of the Stieltjes transform of eq. (1.63), using arguments very similar to the cavity method of statistical physics. These different approaches are detailed in [AGZ10], and we will apply the Stieltjes transform method to other kinds of random matrices in Chapter 5 of this thesis. Finally, we will also present in Section 1.5.2 a third way to prove Theorem 1.6, using a *Coulomb Gas* representation of the eigenvalues. As we will see, this method grants access not only to the convergence of the ESD but also to its large deviations away from its expected value.

Semicircle law and free probability – Intuitively, the semicircle law plays in free probability a role analogous to the one of the normal distribution in classical probability. Indeed its free cumulants c_k (cf. eq (1.65)) satisfy $c_k = 0$ for $k \geq 3$, much like the cumulants of the Gaussian distribution are zero for orders greater or equal to 3.

A second very natural class of random matrices to study are the Wishart matrices of eq. (1.62). As we mentioned, this was done first in [MP67], while some generalizations and precisions were brought after, e.g. in [SB95]. Note first that the matrix of eq. (1.62) has the same spectrum, up to zeros, as $\mathbf{Z}^\dagger \mathbf{\Sigma} \mathbf{Z}/m$ (with $\mathbf{Z} \in \mathbb{R}^{n \times m}$ an i.i.d. Gaussian matrix) so that we will study this matrix in the following. We state here the main result of [MP67] in the form of Theorem 1.1 of [SB95], naming it the *Marchenko-Pastur equation*.

Theorem 1.7 (*Marchenko-Pastur equation [MP67]*)

Let $p, k \to \infty$ with $p/k \to \alpha > 0$. Let $\mathbf{W} \in \mathbb{K}^{p \times k}$ a matrix whose elements are drawn i.i.d. from $\mathcal{N}_\beta(0,1)$. Let $\mathbf{T}_p \in \mathcal{H}_p(\mathbb{K})$ be a random Hermitian matrix, independent of $\mathbf{W}$, such that the ESD of $\mathbf{T}_p$ converges weakly (and a.s.) to a measure ν_T. Then, almost surely, the ESD of $\mathbf{B}_k \equiv \mathbf{W}^\dagger \mathbf{T}_p \mathbf{W}/k$ converges in law to $\mu_B \in \mathcal{M}_1^+(\mathbb{R})$, whose Stieltjes transform satisfies, for every $z \in \mathbb{C}_+$:

$$\mathcal{S}_{\mu_B}(z) = -\Big[z - \alpha \int \nu_T(\mathrm{d}t) \frac{t}{1 + t\mathcal{S}_{\mu_B}(z)}\Big]^{-1}. \tag{1.68}$$

Moreover, for every $z \in \mathbb{C}_+$, there is a unique solution to eq. (1.68) such that $\mathcal{S}_{\mu_B}(z) \in \mathbb{C}_+$. This equation thus characterizes uniquely the measure μ_B.

Theorem 1.7 is proven by means of the cavity method. We refer the reader to the random matrix proofs of Chapter 5, where we will re-derive Theorem 1.7. Note that eq. (1.68) can be written

as an explicit formula in terms of the R-transform of eq. (1.64):

$$\mathcal{R}_{\mu_B}(s) = \alpha \int \nu_T(\mathrm{d}t) \frac{t}{1-ts}.$$

The Marchenko-Pastur equation has been generalized further, e.g. by considering non-linear transformations of the elements of a Wishart matrix [BP19, PW19, PS21]. To conclude, let us mention the most celebrated application case of Theorem 1.7, when $\mathbf{T}_p = \mathrm{I}_p$. This corresponds to Wishart matrices with identity covariance, and in this case the LSD is called the *Marchenko-Pastur distribution* $\mu_B = \mu_{\mathrm{MP},\alpha}$. Its density $\rho_{\mathrm{MP},\alpha}$ is given by:

$$\rho_{\mathrm{MP},\alpha}(x) = \max(0, 1-\alpha)\delta(x) + \frac{\sqrt{(\lambda_+(\alpha)-x)(x-\lambda_-(\alpha))}}{2\pi x} \mathbb{1}\{\lambda_-(\alpha) < x < \lambda_+(\alpha)\}, \tag{1.69}$$

with $\lambda_\pm(\alpha) \equiv (1 \pm \sqrt{\alpha})^2$.

1.5.2 Large deviations

Large deviations theory is a set of probabilistic techniques, heavily used in mathematical physics, that focuses on the study of extremely rare events. A large part of statistical mechanics can actually be understood as a consequence of large deviations properties.

References – Our description of large deviations theory and its results will be extremely far from complete, as we simply aim at introducing a few important definitions and theorems. We refer the reader to e.g. [DZ98, Kle13] for more complete mathematical discussions.

To introduce large deviation principles, let us consider a sequence $(X_i)_{i=1}^n$ of i.i.d. random variables, with common mean m and variance σ^2. By the strong law of large numbers we know

$$S_n \equiv \frac{1}{n}\sum_{i=1}^n X_i \underset{n\to\infty}{\overset{\text{a.s.}}{\to}} m. \tag{1.70}$$

The goal of large deviations theory is to understand the speed of this convergence, which is equivalent to understand large deviations of S_n from its limit value. Note that while the central limit theorem gives that $\sqrt{n}(S_n - m) \underset{n\to\infty}{\overset{\text{weakly}}{\to}} \mathcal{N}(0, \sigma^2)$, large deviations theory focuses on very rare events in which the deviation $S_n - m$ is of order $\mathcal{O}(1)$ rather than in the scale $n^{-1/2}$. For all $x \geq m$ and $t \geq 0$ we can write using Chernoff's inequality that

$$\frac{1}{n}\ln \mathbb{P}[S_n \geq x] \leq -tx + \frac{1}{n}\ln \mathbb{E}\Big[e^{\sum_{i=1}^n tX_i}\Big] \leq -tx + \ln \mathbb{E}\big[e^{tX}\big].$$

So we reach easily that

$$\frac{1}{n}\ln \mathbb{P}[S_n \geq x] \leq -\sup_{t\in\mathbb{R}} \Big\{tx - \ln \mathbb{E}\big[e^{tX}\big]\Big\}. \tag{1.71}$$

Actually, as first shown by Cramér [Cra38], the bound of eq. (1.71) is tight as $n \to \infty$. The perhaps easiest way to prove this is to use a *tilting* of the measure, a technique that we will use again in Chapter 8 in a random matrix context, and the reader can refer to e.g. [Kle13] for a detailed proof. In the end, we reach:

$$\lim_{n\to\infty} \frac{1}{n}\ln \mathbb{P}[S_n \geq x] = -\sup_{t\in\mathbb{R}} \Big\{tx - \ln \mathbb{E}\big[e^{tX}\big]\Big\}. \tag{1.72}$$

Eq. (1.72) exactly describes what we call a *large deviation principle*, here in the scale n: it details the exponential tail of the probability of a random variable. It motivates the following general definition, stated here in a precise mathematical language.

Definition 1.4 (*Large deviations of real variables*)

Let $\{\mu_n\}_{n\geq 1}$ be a sequence of Borel probability measures on $\mathbb{R}$ and $\{a_n\}_{n\geq 1}$ a sequence of positive reals such that $\lim_{n\to\infty} a_n = +\infty$. The sequence $\{\mu_n\}_{n\geq 1}$ satisfies a *large deviation principle* (LDP) with speed $\{a_n\}$ if for any Borel measurable set $E \subseteq \mathbb{R}$[14]:

$$-\inf_{x\in E^\circ} I(x) \leq \liminf_{n\to\infty} \frac{1}{a_n} \ln \mu_n(E) \leq \limsup_{n\to\infty} \frac{1}{a_n} \ln \mu_n(E) \leq -\inf_{x\in \overline{E}} I(x), \tag{1.73}$$

with a lower semi-continuous $I : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ called the *rate function*.

Note in particular that if the random variables concentrate, as $n \to \infty$, to a limit value m (as in eq. (1.70)), then $I(m) = 0$ is the global minimum of the rate function.

The physics language – In theoretical physics, LDPs are usually stated in a much simpler, informal, way. We consider a sequence P_n of PDFs, and we state the LDP of Definition 1.4 as

$$\frac{1}{n} \ln P_n(x) \simeq -I(x), \qquad \text{or} \qquad P_n(x) \simeq e^{-nI(x)}. \tag{1.74}$$

Eq. (1.72) can then be turned into what is known as Cramér's theorem:

Theorem 1.8 (*Cramér's theorem, [Cra38]*)

Let $(X_i)_{i=1}^n$ be a sequence of i.i.d. real random variables such that $\Lambda(t) \equiv \ln \mathbb{E}[\exp(tX)]$ is finite for all $t \in \mathbb{R}$. Let $\Lambda^*(u) \equiv \sup_{t\in\mathbb{R}}[tu - \Lambda(t)]$ its Legendre transform. Then the law of the random variables $S_n \equiv (1/n)\sum_{i=1}^n X_i$ satisfies a large deviation principle, in the scale n, with rate function Λ^*.

Another seminal large deviations result, that will prove useful in Chapter 7 of this thesis, is the LDP for the empirical measure of a set of i.i.d. variables, known as *Sanov's theorem*:

Theorem 1.9 (*Sanov's theorem, [San58]*)

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. real random variables with common law μ. We let $\nu_n \equiv n^{-1}\sum_{i=1}^n \delta_{X_i}$ the (random) empirical measures of the sequence. Then the law of ν_n satisfies, as $n \to \infty$, a large deviation principle in the scale n, with rate function given by the Kullback-Leibler divergence[15]:

$$I(\nu) \equiv D_{\mathrm{KL}}(\nu|\mu) = \begin{cases} \displaystyle\int \mathrm{d}\nu \ln \frac{\mathrm{d}\nu}{\mathrm{d}\mu} & \text{if } \nu \ll \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

One of the most important applications of large deviation principles is the ability to use them to compute the exponential limit of some high-dimensional integrals, by use of *Varadhan's lemma* (named after one of the most prolific mathematicians in large deviations theory, who obtained in 2007 the Abel Prize for its contributions to the field):

[14] Recall that E° and $\overline{E}$ are the interior and adherence of E.

[15] Recall that $\nu \ll \mu$ means absolute continuity of ν with respect to μ, i.e. for any measurable set A we have $\mu(A) = 0 \Rightarrow \nu(A) = 0$. By the Radon-Nikodym theorem [Nik30], we know that $\nu \ll \mu$ is equivalent to the existence of the Radon-Nikodym derivative $\mathrm{d}\nu/\mathrm{d}\mu$ such that $\nu(A) = \int_A (\mathrm{d}\nu/\mathrm{d}\mu)\mathrm{d}\mu$.

Lemma 1.10 (*Varadhan's lemma, [DZ98]*)

Let $(X_n)_{n\geq 1}$ be a sequence of real random variables that satisfies a large deviation principle with rate function $I(x)$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous function, and assume that one of the two following conditions hold:

$$\begin{cases} \lim\limits_{A\to\infty} \limsup\limits_{n\to\infty} \dfrac{1}{n} \ln \mathbb{E}\big[\exp\{n\,\varphi(X_n)\}\mathbb{1}[\phi(X_n) \geq A]\big] = -\infty, \quad \text{or} \\ \limsup\limits_{n\to\infty} \dfrac{1}{n} \ln \mathbb{E}\big[\exp\{n\,\gamma\,\varphi(X_n)\}\big] < \infty \qquad \text{for some } \gamma > 1. \end{cases}$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\big[\exp\{n\,\varphi(X_n)\}\big] = \sup_{x\in\mathbb{R}} [\varphi(x) - I(x)]. \tag{1.75}$$

How physicists say it – Note that theoretical physicists usually do not refer to Varadhan's lemma. As the large deviation principle is stated in the form of eq. (1.74), we can write:

$$\frac{1}{n} \ln \int \mathrm{d}x P_n(x)\, e^{n\varphi(x)} \simeq \frac{1}{n} \ln \int \mathrm{d}x\, e^{n(\varphi(x)-I(x))} \simeq \sup_{x\in\mathbb{R}} [\varphi(x) - I(x)], \tag{1.76}$$

which is usually mentioned as an application of Laplace's method.

Application I: the Gibbs-Boltzmann measure and the canonical ensemble – Sanov's theorem 1.9 is intimately related to asymptotic quantities described in statistical mechanics. It formalizes one of the most important physical properties of statistical systems: when given a set of thermodynamical constraints, the distribution of a system will maximize the entropy while satisfying these constraints. Indeed, let us consider a set of n i.i.d. real random variables $\{X_i\}$ (representing e.g. the position or velocity of particles), with a common law μ. Directly applying Sanov's theorem then tells us about the *microcanonical* description of the ensemble: as $n \to \infty$, the distribution of the particles will concentrate on the distribution ν^* that maximizes the entropy, i.e. that minimizes $D_{\mathrm{KL}}(\nu|\mu)$. Trivially, this distribution is $\nu^* = \mu$, and this result is intimately related to the fundamental postulate of statistical physics: in an isolated system at equilibrium, all microstates are equally probable.

Let us now assume to have access to an (intensive) *energy* function $\mathcal{E}(\{X_i\}) = n^{-1} \sum_{i=1}^n E(X_i)$, which is thus a function of the empirical measure $\nu_n \equiv n^{-1} \sum_{i=1}^n \delta_{X_i}$ (e.g. the kinetic energy $E_{\mathrm{kin}} = (m/2) \sum_i v_i^2$ if the variables X_i represent the velocity of particles). We want to study the distribution of the $\{X_i\}$ for a *fixed value $\mathcal{E}$ of the total intensive energy*. By Sanov's theorem 1.9, we know that the asymptotic distribution of $\{X_i\}$ will concentrate as $n \to \infty$ to the maximum-entropy distribution satisfying the constraint:

$$\nu(\mathcal{E}) \equiv \underset{\substack{\nu\in\mathcal{M}_1^+(\mathbb{R}) \\ \text{s.t. } \int \nu(\mathrm{d}x)E(x)=\mathcal{E}}}{\arg\min} [D_{\mathrm{KL}}(\nu|\mu)]. \tag{1.77}$$

The solution to this variational problem is quite easy. We introduce a Lagrange multiplier η (our choice of notation will become clear very soon) to enforce the constraint, and we reach that

$$\nu(\mathcal{E}) = \nu_\eta(\mathcal{E}) \equiv \underset{\nu\in\mathcal{M}_1^+(\mathbb{R})}{\arg\min} [\eta \int \nu(\mathrm{d}x)E(x) + D_{\mathrm{KL}}(\nu|\mu)], \tag{1.78}$$

with $\eta = \eta(\mathcal{E})$ chosen such that $\int \nu_\eta(\mathcal{E})(\mathrm{d}x)E(x) = \mathcal{E}$. It is straightforward to solve eq. (1.78) and we reach:

$$\nu(\mathcal{E})(\mathrm{d}x) = \frac{\mu(\mathrm{d}x)e^{-\eta E(x)}}{\mathcal{Z}}, \tag{1.79}$$

with $\mathcal{Z}$ chosen to ensure normalization of the distribution. Eq. (1.79) is precisely the *Gibbs-Boltzmann* measure of statistical physics we introduced in Section 1.1.4: it is the distribution that maximizes entropy at a fixed energy level. In statistical physics, the *canonical* ensemble consists in studying the dual problem: we fix a $\eta > 0$ (usually called *inverse temperature*), and we study the distribution of eq. (1.79). This description is dual (and thus equivalent) to the path we took here, with the energy given by $\mathcal{E} = \mathcal{Z}^{-1} \int \mu(\mathrm{d}x)\, E(x)\, e^{-\eta E(x)}$.

Application II: large deviations and the semicircle law – Large deviations are well-suited to describe asymptotic phenomena that frequently appear in the physics literature. A good example is the derivation of Theorem 1.6 when the entries of the Wigner matrix are i.i.d. Gaussian (real or complex). Indeed, let us consider a Gaussian Wigner matrix $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$. In this case, the eigenvalues $\{\lambda_i\}_{i=1}^n$ of $\mathbf{J}$ are independent of the eigenvectors (which are Haar distributed). It is then easy to see by a change of variables of the original Gaussian matrix measure to the eigenvalues that (see e.g. Proposition 4.1.1 of [AGZ10]):

$$\mathbb{P}(\mathrm{d}\lambda_1, \cdots, \mathrm{d}\lambda_n) = \frac{1}{\mathcal{Z}_n} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-\frac{\beta n}{4}\sum_{i=1}^n \lambda_i^2} \prod_{i=1}^n \mathrm{d}\lambda_i. \tag{1.80}$$

As we saw in Section 1.5.1, a particularly important quantity to study in random matrix theory is the empirical spectral distribution (ESD) $\nu_n \equiv n^{-1}\sum_{i=1}^n \delta_{\lambda_i}$. Eq. (1.80) can be interpreted as the Gibbs-Boltzmann measure (at temperature $T = 1$) of an interacting gas of particles, subject to a logarithmic repulsion potential and a confining parabolic potential, resulting in the energy:

$$\frac{1}{n^2}E(\{\lambda_i\}) \equiv \frac{\beta}{4}\int \nu_n(\mathrm{d}x)\, x^2 - \frac{\beta}{2}\iint_{x\neq y} \nu_n(\mathrm{d}x)\nu_n(\mathrm{d}y)\ln|x-y|.$$

Because of the logarithmic repulsion potential, this representation is usually called in the physics literature a *Coulomb gas*. Let us wear a physicist's hat for a moment to state an informal argument that will allow us to derive the large deviations of ν_n from the previous description. As we know from Sanov's theorem 1.9, we can write, with $\rho_n(x)$ the density of ν_n:

$$\int \prod_{i=1}^n \mathrm{d}\lambda_i\, \delta\Big[\rho_n(x) - \frac{1}{n}\sum_{i=1}^n \delta(x-\lambda_i)\Big] \simeq \exp\Big\{-n\int \mathrm{d}x\, \rho_n(x)\ln\rho_n(x)\Big\}. \tag{1.81}$$

Therefore we reach from eqs. (1.80) and (1.81):

$$P[\nu_n] \simeq \frac{1}{\mathcal{Z}_n} e^{-n^2\left[\frac{\beta}{4}\int \mathrm{d}x\, \rho_n(x)\, x^2 - \frac{\beta}{2}\iint_{x\neq y}\mathrm{d}x\,\mathrm{d}y\,\rho_n(x)\rho_n(y)\ln|x-y|\right] - n\int \mathrm{d}x\,\rho_n(x)\ln\rho_n(x)}. \tag{1.82}$$

Note that as the large deviations described by Sanov's theorem 1.9 are in the scale n while the energy $E(\{\lambda_i\})$ scales as n^2, this latter term will be the only contribution to the large deviations at leading order! In physics terms, the free energy of this Coulomb gas is dominated solely by the energetic component at leading order, while we can discard the entropic contribution. This informal argument derived from eq. (1.82) leads to the conjecture thaw the law of ν_n satisfies a large deviation principle, in the scale n^2, with rate function:

$$I(\nu) \equiv -\frac{\beta}{4}\int \nu(\mathrm{d}x)\, x^2 + \frac{\beta}{2}\iint \nu(\mathrm{d}x)\nu(\mathrm{d}y)\ln|x-y| - C, \tag{1.83}$$

with C chosen such that $\inf_{\nu\in\mathcal{M}_1^+(\mathbb{R})} I(\nu) = 0$[16]. This statement is indeed rigorous, and its proof is done in [BAG97]. By a straightforward computation left to the reader (done e.g. in [LNV18], or in [BAG97]), one can check that

$$\sigma_{\text{s.c.}} = \underset{\nu\in\mathcal{M}_1^+(\mathbb{R})}{\arg\min}\, I(\nu), \quad \text{and} \quad I(\sigma_{\text{s.c.}}) = 0, \tag{1.84}$$

and that this minimum is unique. Therefore we have re-derived Theorem 1.6 as a simple corollary of our large deviations result! Our result is actually much stronger, as we showed that the convergence of the spectral measure happens in the scale $e^{\Theta(n^2)}$.

1.5.3 High-dimensional "spherical" integrals

In this section we introduce quantities known as *spherical integrals*, which can be seen as the equivalent of Laplace/Fourier transforms in the context of random matrices. Precisely, we consider integrals of the type ($\mathcal{D}\mathbf{O}$ is the Haar measure on the compact group $\mathcal{U}_\beta(n)$)

$$I_n(\mathbf{A},\mathbf{B}) \equiv \int_{\mathcal{U}_\beta(n)} \mathcal{D}\mathbf{O}\, \exp\{n\text{Tr}[\mathbf{A}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger]\}, \tag{1.85}$$

in which $\mathbf{A},\mathbf{B}$ are Hermitian/symmetric random matrices, and we are interested in the large-n behavior of I_n. Applications of high-dimensional spherical integrals in statistical physics and random matrix theory are numerous. They have been studied in the context of 2-spin glass models [KTJ76, MPR94b, PP95], and they allow e.g. to derive the density of the eigenvalue distribution of random matrices [Zub18, CMZ19] or the large deviations of the eigenvalues [BG20, GH20, Hus20, BGH20, AGH21, McK21b, GH21] (see also Chapter 8 of this thesis).

Closed formula at any n – For the unitary group, Harish-Chandra [HC57], followed by Itzykson and Zuber [IZ80], derived explicit formulas for these integrals, valid for any dimension[17]. For this reason we will generally refer to integrals of the type of eq. (1.85) as *Harish-Chandra-Itzykson-Zuber* (HCIZ) integrals. However these formulas are quite involved functions of determinants, and can not easily be used to compute the high-dimensional asymptotics.

We now present results on the high-dimensional limit of different classes of HCIZ integrals. We give references to the literature in which these statements were derived or proven.

Rank-one HCIZ integrals

We start by the simplest case in which one of the two matrices $\mathbf{A},\mathbf{B}$ of eq. (1.85) has rank one. This class of spherical integrals will be important in different approaches taken in this thesis, especially in Chapters 2 and 8.

Theorem 1.11 (*Rank-one HCIZ integral [GM05]*)

Let $\mathbf{J}\in\mathcal{H}_n(\mathbb{K})$ be a matrix such that the Empirical Spectral Distribution (ESD) of $\mathbf{J}$: $\rho_n \equiv n^{-1}\sum_{i=1}^n \delta_{\lambda_i(\mathbf{J})}$ converges a.s. (in the weak sense) as $n\to\infty$ to a distribution ρ with compact support. We also assume that the largest eigenvalue $\lambda_{\max}(\mathbf{J})$ converges a.s. as $n\to\infty$ to $x\in\mathbb{R}$. Let $\theta\geq 0$, and μ_n be the uniform measure on $\mathbb{S}_\beta^{n-1}$. Then

$$I(\theta) \equiv \lim_{n\to\infty}\frac{1}{n}\ln\int \mu_n(\mathrm{d}\mathbf{e})\exp\left\{\frac{\beta\theta n}{2}\mathbf{e}^\dagger\mathbf{J}\mathbf{e}\right\} = \frac{\beta}{2}\inf_{\gamma\geq\theta x}\left[\gamma - \int\rho(\mathrm{d}\lambda)\ln(\gamma-\theta\lambda)\right] - \frac{\beta}{2}. \tag{1.86}$$

[16] This constant comes from the normalization factor $\mathcal{Z}_n$ and ensures that the probability of the limit spectral density approaches 1 as $n\to\infty$.

[17] Note that such an explicit formula does not exist in the real orthogonal case.

Proof of Theorem 1.11 – We give here a simple heuristic derivation of the theorem. Its mathematical proof follows essentially the same lines, and can be found in [GM05]. Let us denote $I_n(\theta)$ the LHS of eq. (1.86) before taking the limit $n \to \infty$. We renormalize it as:

$$I_n(\theta) = \frac{1}{n} \ln \int \mu_n(\mathrm{d}\mathbf{e}) \exp\Big\{\frac{\beta\theta n}{2} \mathbf{e}^\dagger \mathbf{J}\mathbf{e}\Big\} = \frac{1}{n} \ln \frac{\int \mathrm{d}\mathbf{e}\, \delta(\|\mathbf{e}\|^2 - n) \exp\Big\{\frac{\beta\theta}{2} \mathbf{e}^\dagger \mathbf{J}\mathbf{e}\Big\}}{\int \mathrm{d}\mathbf{e}\, \delta(\|\mathbf{e}\|^2 - n)}.$$

We can then introduce a Lagrange multiplier γ to enforce the condition $\|\mathbf{e}\|^2 = n$. At leading order, one obtains:

$$\begin{aligned} I_n(\theta) &= \frac{\frac{1}{n} \ln \int \mathrm{d}\gamma \int_{\mathbb{K}^n} \mathrm{d}\mathbf{x}\; e^{\frac{\beta\theta}{2}\mathbf{x}^\dagger \mathbf{J}\mathbf{x} + \frac{\beta\gamma}{2}(n - \sum_i |x_i|^2)}}{\frac{1}{n} \ln \int \mathrm{d}\gamma \int_{\mathbb{K}^n} \mathrm{d}\mathbf{x}\; e^{\frac{\beta\gamma}{2}(n - \sum_i |x_i|^2)}} = \inf_\gamma \left\{ \frac{1}{n} \ln \frac{\int \mathrm{d}\mathbf{x}\; e^{\frac{\beta\theta}{2}\mathbf{x}^\dagger \mathbf{J}\mathbf{x} + \frac{\beta\gamma}{2}(n - \sum_i |x_i|^2)}}{\exp\{\frac{\beta n}{2}(1 + \ln 2\pi))\}} \right\}, \\ &= \inf_\gamma \Big[\frac{\beta\gamma}{2} - \frac{\beta}{2n} \ln\det\left(\gamma \mathrm{I}_n - \theta \mathbf{J}\right) \Big] - \frac{\beta}{2}. \end{aligned} \tag{1.87}$$

Indeed, one checks easily that the extremum over the Lagrange multiplier γ actually corresponds to an infimum, and this infimum is made over all γ such that $\gamma \mathrm{I}_n - \theta \mathbf{J}$ is positive definite, i.e. $\gamma > \theta \lambda_{\max}(\mathbf{J})$. Since $\lambda_{\max}(\mathbf{J}) \to x$ by hypothesis, we obtain from the $n \to \infty$ limit of eq. (1.87):

$$I(\theta) \equiv \lim_{n\to\infty} I_n(\theta) = \inf_{\gamma > \theta x} \Big[\frac{\beta\gamma}{2} - \frac{\beta}{2} \int \rho(\mathrm{d}\lambda) \ln(\gamma - \theta\lambda) \Big] - \frac{\beta}{2}.$$

Recall indeed that ρ is the LSD of $\mathbf{J}$. This ends the proof. □

Phase transition in rank-one HCIZ integrals

Let us make a few remarks on the form of Theorem 1.11, which will be particularly useful in Chapter 8 when applying these results to derive large deviations of the eigenvalues of random matrices. We denote $\gamma(\theta)$ the solution to the infimum in eq. (1.86). If this infimum arises for $\gamma(\theta) > \theta x$, then it satisfies the saddle-point equation

$$\int \frac{\rho(\mathrm{d}\lambda)}{\gamma(\theta) - \theta\lambda} = 1. \tag{1.88}$$

Eq. (1.88) then has the solution $\gamma(\theta) = \theta \mathcal{S}_\rho^{-1}(-\theta)$, as long as $-\mathcal{S}_\rho(x) \geq \theta$, where $\mathcal{S}_\rho$ is the Stieltjes transform of ρ, introduced in Section 1.5.1[18]. If rather $\theta > -\mathcal{S}_\rho(x)$ then γ "sticks" to the solution $\gamma = \theta x$. In the end, we can compute $I_n(\theta)$ in the small-θ (or *high-temperature* in the physics language) phase $\theta \leq \theta_c(x) \equiv -\mathcal{S}_\rho(x)$:

$$I(\theta) = \frac{\beta\theta}{2} \mathcal{R}_\rho(\theta) - \frac{\beta}{2} \int \rho(\mathrm{d}\lambda) \ln[1 + \theta \mathcal{R}_\rho(\theta) - \theta\lambda],$$

in which $\mathcal{R}_\rho(x) \equiv \mathcal{S}_\rho^{-1}(-x) - x^{-1}$ is the R-transform of ρ, cf. Section 1.5.1. By taking the derivative of this last expression with respect to θ it is easy to show that it simplifies to:

$$I(\theta) = \frac{\beta}{2} \int_0^\theta \mathcal{R}_\rho(u) \mathrm{d}u. \tag{1.89}$$

In the "low-temperature" phase $\theta \geq \theta_c(x) = -\mathcal{S}_\rho(x)$ one rather has

$$I(\theta) = \frac{\beta}{2} \Big[-1 + \theta x - \ln\theta - \int \rho(\mathrm{d}\lambda) \ln(x - \lambda) \Big]. \tag{1.90}$$

[18] If x is the right edge of the bulk of ρ, we denote $S_\rho(x) \equiv \lim_{z \downarrow x} S_\rho(z)$.

Note that in both phases it is clear that $I(\theta)$ is concentrating (or "self-averaging" in the physics language), as it only depends on $\mathbf{J}$ via its ESD and its largest eigenvalue, which we assumed converge a.s. Moreover its series expansion around $\theta = 0$ is related to the free cumulants of ρ, cf. eq. (1.65):

$$I(\theta) = \frac{\beta}{2} \sum_{p=1}^{\infty} \frac{c_p(\rho)}{p} \theta^p, \tag{1.91}$$

which is valid in the high-temperature phase $\theta \leq \theta_c(x)$.

Finite-rank $k \geq 1$

Let us now assume that one of the two matrices $\mathbf{A}, \mathbf{B}$ in eq. (1.85) has finite rank $k \geq 1$ as $n \to \infty$. Then one can state very natural generalizations of the previous results to this setting. Informally, the integral in this case behaves at leading order as k decoupled rank-one integrals described by Theorem 1.11, which is shown in detail in the following theorem.

Theorem 1.12 (*Finite-rank HCIZ integral [GM05, CŚ07]*)

Let $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ as in Theorem 1.11. Let $\theta_1, \cdots, \theta_k \geq 0$, and $\mathbf{e}_1, \cdots, \mathbf{e}_k$ be the first k columns of a Haar-distributed matrix $\mathbf{O} \in \mathcal{U}_\beta(n)$. Then

$$I_k(\theta_1, \cdots, \theta_k) \equiv \lim_{n\to\infty} \frac{1}{nk} \ln \mathbb{E}\Big[\exp\Big\{ \sum_{a=1}^{k} \frac{\beta\theta_a n}{2} \mathbf{e}_a^\dagger \mathbf{J} \mathbf{e}_a \Big\}\Big] = \frac{1}{k} \sum_{a=1}^{k} I_1(\theta_a), \tag{1.92}$$

in which the expectation is done over $\{\mathbf{e}_1, \cdots, \mathbf{e}_k\}$ and I_1 is the asymptotic of the rank-one case given in Theorem 1.11.

Rectangular spherical integrals

Interestingly, the asymptotic results we described so far on spherical integrals can be generalized to a class of "rectangular" spherical integrals. We state this generalization for rank-one integrals, but it can be written for finite-rank integrals along the same lines we described for usual spherical integrals. These rectangular HCIZ integrals will be useful at several points in this thesis, in very different contexts, e.g. in Chapters 2, 6 and 8.

Theorem 1.13 (*Rank-one "rectangular" HCIZ integral [Kab08a, Kab08b, BG11]*)

Let $\mathbf{L} \in \mathbb{K}^{m\times n}$, such that $\mathbf{J} \equiv \mathbf{L}^\dagger \mathbf{L}$ satisfies hypothesis (ii) of Model S. We assume that $n, m \to \infty$ with $m/n \to \alpha > 0$. Assume moreover that $\lambda_{\max}(\mathbf{J})$ converges a.s. as $n \to \infty$ to $x \geq 0$. Let $\theta \geq 0$, and μ_n be the uniform measure on $\mathbb{S}_\beta^{n-1}$. Then

$$\begin{aligned} I_{\text{rect.}}(\theta) &\equiv \lim_{n\to\infty} \frac{1}{n} \ln \int \mu_m(\mathrm{d}\mathbf{e}) \mu_n(\mathrm{d}\mathbf{f}) \exp\Big\{ \beta\sqrt{\alpha}\theta n (\mathbf{e}^\dagger \mathbf{L} \mathbf{f}) \Big\} \\ &= \frac{\beta}{2} \Big\{ \inf_{\substack{\gamma_e, \gamma_f > 0 \\ \gamma_e\gamma_f \geq \theta^2 x}} \Big[\gamma_f + \alpha\gamma_e - (\alpha - 1) \ln \gamma_e - \int \rho(\mathrm{d}\lambda) \ln(\gamma_e\gamma_f - \theta^2\lambda) \Big] - \frac{1+\alpha}{2} \Big\}. \end{aligned}$$

The function $I_{\text{rect.}}(\theta)$ is closely related to the *rectangular R-transform* introduced in [BG11]. While we do not detail the definitions of this function, in analogy with eq. (1.91) we define a set of coefficients $\Gamma_p(\alpha, \rho)$ (analogous to the free cumulants in this rectangular context) by the

analytical expansion around $\theta = 0$:

$$I_{\text{rect.}}(\theta) = \frac{\beta}{2} \sum_{p=1}^{\infty} \frac{\Gamma_p(\alpha, \rho)}{p} \theta^{2p}. \tag{1.93}$$

As in the symmetric case, the function $I_{\text{rect.}}(\theta)$ might admit phase transitions corresponding to a "saturation" of the Lagrange multipliers γ_e, γ_f at the largest eigenvalue, i.e. when $\gamma_e \gamma_f = \theta^2 x$. Such transitions will be derived and used in Chapter 8.

Diverging rank $k \to \infty$

If one of the two matrices $\mathbf{A}, \mathbf{B}$ has extensive rank in n while the other one has a diverging rank $k = k(n) \to \infty$ as $n \to \infty$, the computations become *a priori* more complicated. While we will not directly use these results in this dissertation, solutions have been found in different regimes of $k(n)$ and the resulting picture is quite simple. Indeed, for $k = \mathcal{O}(n)$, the HCIZ integral behaves as in the finite-rank case: its limit is given by k decoupled rank-one integrals. This was first shown in [GM05] for $k(n) = \mathcal{O}(n^{1/2-\epsilon})$ (for arbitrary $\epsilon > 0$), and later generalized in [CŚ07] to any $k(n) = \mathcal{O}(n)$.

Theorem 1.14 (*Slowly-diverging-rank HCIZ integral [GM05, CŚ07]*)

Let $\mathbf{J} \in \mathcal{H}_n(\mathbb{K})$ as in Theorem 1.11 and $k = k(n)$ such that $\lim_{n\to\infty} k(n) = +\infty$ and $k(n) = \mathcal{O}(n)$. Let $\theta_1, \cdots, \theta_k \geq 0$ such that $k^{-1} \sum_{a=1}^{k} \delta_{\theta_a}$ converges weakly to $\mu \in \mathcal{M}_1^+(\mathbb{R})$, and we assume that there exists $C > 0$ that uniformly bounds $\theta_a \leq C$. We consider $\mathbf{e}_1, \cdots, \mathbf{e}_k$ the first k columns of a Haar-distributed matrix $\mathbf{O} \in \mathcal{U}_\beta(n)$. Then

$$I(\mu) \equiv \lim_{n\to\infty} \frac{1}{nk(n)} \ln \mathbb{E}\Big[\exp\Big\{ \sum_{a=1}^{k(n)} \frac{\beta \theta_a n}{2} \mathbf{e}_a^\dagger \mathbf{J} \mathbf{e}_a \Big\}\Big] = \int \mu(\mathrm{d}\theta) I_1(\theta), \tag{1.94}$$

in which the expectation is done over $\{\mathbf{e}_1, \cdots, \mathbf{e}_k\}$ and I_1 is the asymptotic of the rank-one integral given in Theorem 1.11.

On the other hand the case $k(n) = \Theta(n)$ is much more involved and can not be reduced to the rank-one case. It was first studied by Matytsin [Mat94] and later proven in [GZ02].

Theorem 1.15 (*Extensive-rank HCIZ integral [Mat94, GZ02]*)

Let $n \geq 1$, and $\mathbf{A}, \mathbf{B} \in \mathcal{H}_n(\mathbb{K})$. We assume that the ESDs of $\mathbf{A}$ and $\mathbf{B}$ both converge a.s. as $n \to \infty$ (in the weak sense) to probability measures $\mu_A, \mu_B \in \mathcal{M}_1^+(\mathbb{R})$. Then:

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n^2} \ln \int_{\mathcal{U}_\beta(n)} \mathcal{D}\mathbf{O} \, \exp\Big\{ \frac{\beta n}{2} \mathrm{Tr}[\mathbf{A}\mathbf{O}\mathbf{B}\mathbf{O}^\dagger] \Big\} \\ &= \frac{\beta}{2} \Bigg[-\frac{3}{4} + J(\mu_A) + J(\mu_B) - \frac{1}{2} \inf_{\rho, m} \Big\{ \int_0^1 \int \mathrm{d}t \, \mathrm{d}x \, \Big(\frac{m_t(x)^2}{\rho_t(x)} + \frac{\pi^2}{3} \rho_t(x)^3 \Big) \Big\} \Bigg] \end{aligned}$$

Let us denote the measure-valued process $\mu_t(\mathrm{d}x) \equiv \rho_t(x)\mathrm{d}x$. The infimum in the above equation is done over m, ρ satisfying the Euler continuity equation $\partial_t \rho_t(x) + \partial_x m_t(x) = 0$, and the boundary conditions $\mu_0 = \mu_A$ and $\mu_1 = \mu_B$. This result assumes finally that $J(\mu_A)$ and $J(\mu_B)$ are finite, where for any $\mu \in \mathcal{M}_1^+(\mathbb{R})$ we define:

$$J(\mu) \equiv \frac{1}{2} \int \mu(\mathrm{d}x) \, x^2 - \frac{1}{2} \iint \mu(\mathrm{d}x) \mu(\mathrm{d}y) \ln |x - y|.$$

Chapter 2

Revisiting high-temperature expansions

"One of the principal objects of theoretical research in my department of knowledge is to find the point of view from which the subject appears in its greatest simplicity."

J. Willard Gibbs, Proceedings of the American Academy of Arts and Sciences (1881).

Disclaimer – In this chapter, we revisit one of the oldest theoretical tools of the statistical physics of disordered systems: *high-temperature expansions*, initiated by Plefka [Ple82] and refined by Georges and Yedidia [GY91]. This chapter is based on the published work [MFC+19], and serves both as a presentation of its results and as a pedagogical introduction to high-temperature expansions, which will prove useful in particular to tackle the involved problem of extensive-rank matrix factorization in Chapter 3.

2.1 Organization of the chapter and main results

Beyond i.i.d. couplings – As we saw in Chapter 1, the statistical physics approach, combining the TAP free energy and message-passing algorithms, is especially powerful when the coupling constants in the underlying statistical model are distributed as i.i.d. variables. This is, of course, a strong limitation and many strategies have been designed to improve on it, and we introduced some of them in Section 1.4.3: the adaptive TAP (adaTAP) method [OW01a, OW01b], approximation schemes such as *Expectation Consistency* (EC) [Min01, OW05a] and the recent improvements of AMP such as *Vector Approximate Message Passing* (VAMP) and its many variants [SRF16, OCW16, COFW16, MP17, RSF17]. Given these different strategies, one may wonder when they actually lead to asymptotically exact inference. In this chapter, we address this question using *high-temperature* expansions.

A short history of high-temperature expansions – High-temperature expansions at fixed order parameters, denoted in this thesis as *Plefka* or *Plefka-Georges-Yedidia* (PGY) expansions, have historically been an important tool of the study of disordered systems. In the context of spin glass models they have been introduced by Plefka [Ple82] for the Sherrington-Kirkpatrick (SK) model, and have been subsequently generalized by Georges and Yedidia [GY91]. This latter paper provides a systematic way to compute high-temperature (or high-dimensional) expansions of the TAP free entropy, which is defined *for a fixed value of the order parameters*.

Our goal – One important aim of the present chapter is to apply this method to a general class of inference problems with pairwise interactions, in which the coupling constants are not i.i.d., but rather possess correlations, satisfying a rotational invariance property. This encapsulates many widely studied models, such as Restricted Boltzmann Machines (RBMs) or Generalized Linear Models (GLMs) with correlated data matrices, introduced in Section 1.1. In particular, we generalize earlier and inspirational work by Parisi and Potters [PP95], who computed

the self-consistent equations for the marginals in Ising models with orthogonal couplings via a resummation of the infinite series given by the high-temperature expansion. In our general setting, we shall show that a similar resummation yields the EC, adaTAP and VAMP formalisms: they are all equivalent as we saw in Section 1.4.3, and we conjecture that they are exact in the thermodynamic limit in the replica symmetric phases. On the way to these conclusions we will uncover diagrammatical results in connection with free probability and random matrix theory.

Organization of the chapter and main results –

- **Spherical models with rotationally-invariant couplings** – We first provide a pedagogical introduction to "high-temperature" PGY expansions, inspired by the work of Georges-Yedidia [GY91] for Ising models, adding some new results on the diagrammatics of the expansions. This yields a general framework that encapsulates many known properties of systems sharing this pairwise structure. In the corresponding section 2.2, we focus on spherical models and we generalize the seminal works of [MPR94a, MPR94b, PP95]. While these works studied Ising models with orthogonal couplings, we consider spherical models with more general rotationally invariant couplings. We examine two types of models: "symmetric" models with an interaction of the type $\mathbf{x}^\intercal \mathbf{J}\mathbf{x}$, in which $\mathbf{J}$ follows Model S, and "bipartite" models with interactions of the type $\mathbf{h}^\intercal \mathbf{F}\mathbf{x}$, in which $\mathbf{F}$ follows Model R. This encapsulates orthogonal couplings, but can also be applied to other random matrix ensembles such as the Gaussian Orthogonal Ensemble (GOE), the Wishart ensemble, and many others. Using diagrammatic results that we derive with random matrix theory, we conjecture a resummation of the PGY expansion giving the Gibbs free entropy in these models. Our results are in particular consistent with the findings of classical works for Gaussian couplings [Ple82] and orthogonal couplings [PP95].
- **PGY expansion for statistical models with correlated couplings** – Section 2.3.1 is devoted to the description of the Plefka expansion for different statistical models and inference problems which possess a coupling or data matrix that has rotation invariance properties. We consider e.g. models similar to the spherical models of Section 2.2, but with generic prior distributions on the underlying variables. Our main conjecture for this part can be stated as the following:

  Conjecture 2.1 (*Informal*)

  For statistical models of symmetric or bipartite interactions with coupling matrices that satisfy respectively Model S or Model R (left and right rotation-invariance), the three equivalent approximations introduced in Section 1.4.3: Expectation Consistency, adaptive TAP and Vector Approximate Message Passing (generically referred to as *EC approximations*), are exact in the large size limit in the high temperature phase.

  We believe that the validity of the above conjecture extends beyond the high temperature phase. In particular that it is correct for inference problems in the Bayes-optimal setting, and more generally anytime the system is in a replica symmetric phase as defined in Chapter 1.

  The approximation behind EC approximations can be checked order by order using our PGY expansion and its resummation. Using diagrammatic results, we show that the EC approximations are exact for these models in the large size limit. We then apply our generic results to different situations, e.g. the Hopfield model [Hop82], compressive sensing, and a very broad class of bipartite models, which includes the Generalized Linear Models (GLMs) with correlated data matrices. We emphasize that we are able to derive the free entropy of all these models using very generic arguments relying only on the rotational invariance of the problem.
- **The TAP equations and message-passing algorithms** – In Section 2.3.2, we show that the TAP equations that we derived by PGY expansion in rotationally invariant models can be understood as the fixed point equations of message-passing algorithms. In the converse

way, many message-passing algorithms can be seen as an iteration scheme of these TAP equations. While this was known in several models in which the underlying data matrix was assumed to be i.i.d. (cf. Chapter 1), using our diagrammatic results we are able to generalize these correspondences to correlated models. We argue that the stationary limit of the Vector Approximate Message Passing (VAMP) algorithm [RSF17] for compressed sensing with correlated matrices gives back our TAP equations derived via PGY expansion. Even more generally, the Generalized Vector Approximate Passing (G-VAMP) algorithm [SRF16], defined for the very broad class of Generalized Linear Models with correlated matrices, yields fixed point equations that are equivalent to our PGY-expanded TAP equations.

- **Diagrammatics of the expansion** – Our results are largely based on a precise control on the diagrammatics of the PGY expansion for rotationally invariant matrices, which are presented in Section 2.4. In particular, we leverage mathematical results on HCIZ integrals (cf. Section 1.5.3) to argue that only a very specific class of diagrams contributes to the expansion.

Some technicalities or generalizations will be deferred to Appendix A.

Additional notation – In this chapter we will sometimes write $A_n \simeq B_n$ to denote that

$$\frac{1}{n}\ln A_n = \frac{1}{n}\ln B_n + \mathcal{O}_n(1).$$

2.2 Plefka-Georges-Yedidia expansion step-by-step

2.2.1 Pedagogical derivation for a spherical SK-like model

In this section $n \geq 1$, $\sigma > 0$, and we define the following pairwise interaction Hamiltonian on $\mathbb{S}^{n-1}(\sigma\sqrt{n})$, the n-th dimensional sphere of radius $\sigma\sqrt{n}$:

$$H_{\mathbf{J}}(\mathbf{x}) \equiv -\frac{1}{2}\mathbf{x}^\intercal \mathbf{J}\mathbf{x} = -\frac{1}{2}\sum_{1\leq i,j\leq n} J_{ij}x_i x_j, \qquad \mathbf{x} \in \mathbb{S}^{n-1}(\sigma\sqrt{n}). \tag{2.1}$$

The coupling matrix $\mathbf{J}$ is a $n\times n$ symmetric random matrix, assumed to be rotationally invariant in the sense of Model S. The interest of this simple "toy" model (which is a spherical "SK-like" model) is that, as we will see, its free energy can be computed exactly so that we can easily control the steps of the expansion by comparison with the exact solution.

Direct free entropy computation

The Gibbs measure associated to the energy of eq. (2.1) at inverse temperature $\eta \geq 0$, and the corresponding free entropy, are[1]:

$$G_{\eta,\mathbf{J}}(\mathrm{d}\mathbf{x}) \equiv \frac{1}{Z_{\eta,\mathbf{J}}} e^{\frac{\eta}{2}\sum_{i,j} J_{ij}x_i x_j}\delta(\|\mathbf{x}\|^2 - n\sigma^2)\mathrm{d}\mathbf{x}, \qquad \Phi_{\mathbf{J}}(\eta) \equiv \lim_{n\to\infty}\frac{1}{n}\ln Z_{\eta,\mathbf{J}}. \tag{2.2}$$

The partition function of the model can directly be written as a function of the spherical HCIZ integrals we studied in Section 1.5.3. We use more precisely Theorem 1.12, and the expression of the spherical integral in the two phases, eqs. (1.89) and (1.90). We reach that in the *high*

[1] We used physicists' convention, introducing a delta constraint on $\|\mathbf{x}\|^2$, while mathematicians usually prefer employing the uniform measure on the sphere $\mathbb{S}^{n-1}(\sqrt{\sigma}n)$, cf. e.g. Theorem 1.12. These conventions are equivalent and only differ by an irrelevant global factor in the partition function.

temperature phase $\eta \leq \eta_c \equiv -\sigma^{-2}\mathcal{S}_\rho(\lambda_{\max})$ the free entropy can be expressed as:

$$\Phi_{\mathbf{J}}(\eta) = \frac{1+\ln 2\pi\sigma^2}{2} + \frac{1}{2}\int_0^{\eta\sigma^2} \mathcal{R}_\rho(x)\mathrm{d}x,$$

in which $\mathcal{R}_\rho(x) \equiv \mathcal{S}_\rho^{-1}(-x) - x^{-1}$ is the R-transform of ρ, cf. Section 1.5. In the low temperature phase $\eta \geq \eta_c = -\sigma^{-2}\mathcal{S}_\rho(\lambda_{\max})$ one rather has

$$\Phi_{\mathbf{J}}(\eta) = \frac{1}{2}\Big[\ln 2\pi + \lambda_{\max}\eta\sigma^2 - \ln\eta - \int \rho(\mathrm{d}\lambda)\ln(\lambda_{\max}-\lambda)\Big].$$

Note that in both phases the free entropy can formally be expressed as:

$$\Phi_{\mathbf{J}}(\eta) = \frac{1}{2}\ln 2\pi + \frac{1}{2}\inf_{\gamma\geq\eta\lambda_{\max}}\Big[\gamma\sigma^2 - \int\rho(\mathrm{d}\lambda)\ln(\gamma-\eta\lambda)\Big], \tag{2.3}$$

a formulation which is more compact and easier to implement algorithmically.

Plefka expansion and the Georges-Yedidia formalism

A more generic way to compute the free entropy is to follow the formalism of [GY91] to perform a high-temperature Plefka expansion [Ple82]. The principle is simple: expand the TAP free entropy of the Gibbs measure of eq. (2.2) at low η, in the high temperature phase. More precisely we will compute the free entropy imposing constraints on the first two moments[2] $\langle x_i\rangle_\eta = m_i$ and $\langle x_i^2\rangle_\eta = v_i + m_i^2$. A set of parameters $\{m_i, v_i\}$ will thus determine a free entropy value, and the comparison with the direct calculation will be made by maximizing the free entropy with respect to $\{m_i, v_i\}$. The spherical constraint $\|\mathbf{x}\|^2 = n\sigma^2$ thus becomes:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^n [v_i + m_i^2]. \tag{2.4}$$

All-in-all, we wish to expand at low η the free entropy *at a given realization of the disorder* (also called *single-graph* free entropy)

$$\Phi_{\mathbf{J}}(\eta,\mathbf{m},\mathbf{v}) \equiv \frac{1}{n}\sum_{i=1}^n\Big[\lambda_i m_i + \frac{\gamma_i}{2}(v_i+m_i^2)\Big] + \frac{1}{n}\ln\int e^{-\eta H_{\mathbf{J}}(\mathbf{x}) - \sum_{i=1}^n\left[\lambda_i x_i + \frac{\gamma_i}{2}x_i^2\right]}\mathrm{d}\mathbf{x}, \tag{2.5}$$

in which we implicitly extremize over the Lagrange multipliers $\{\lambda_i\}$ and $\{\gamma_i\}$ (for lightness of the notations we will keep their dependency on η explicit only when needed).

The Georges-Yedidia operator – In order to perform the expansion we introduce the operator U, defined in [GY91]:

$$U(\eta,\mathbf{J}) \equiv H_{\mathbf{J}} - \langle H_{\mathbf{J}}\rangle_\eta + \sum_{i=1}^n \partial_\eta\lambda_i(\eta)(x_i-m_i) + \frac{1}{2}\sum_{i=1}^n\partial_\eta\gamma_i(\eta)[x_i^2 - v_i - m_i^2]. \tag{2.6}$$

Note that by definition we have $\langle U\rangle_\eta = 0$. One can easily check the following relation for any observable O, which is the main property of the operator U:

$$\frac{\partial\langle O\rangle_\eta}{\partial\eta} = \Big\langle\frac{\partial O}{\partial\eta}\Big\rangle_\eta - \langle OU\rangle_\eta.$$

[2] The notation $\langle\cdot\rangle_\eta$ indicates an average over the Gibbs measure at inverse temperature η.

Therefore since the magnetizations $\{m_i\}$ and variances $\{v_i\}$ do not depend on η this implies:

$$\begin{cases} 0 &= \partial_\eta \langle x_i \rangle_\eta = -\langle x_i U \rangle_\eta = -\langle (x_i - m_i) U \rangle_\eta, \\ 0 &= \partial_\eta \langle x_i^2 \rangle_\eta = -\langle x_i^2 U \rangle_\eta = -\langle (x_i^2 - v_i - m_i^2) U \rangle_\eta. \end{cases}$$

Considering the previous results one can easily compute the derivative of U:

$$\frac{\partial U}{\partial \eta} = \langle U^2 \rangle_\eta + \sum_{i=1}^n \partial_\eta^2 \lambda_i(\eta)(x_i - m_i) + \frac{1}{2} \sum_{i=1}^n \partial_\eta^2 \gamma_i(\eta)[x_i^2 - v_i - m_i^2].$$

Equipped with all the previous relations we can compute the successive derivatives of the TAP free entropy $\Phi_{\mathbf{J}}(\eta, \mathbf{m}, \mathbf{v})$. The reader can easily check that for any $\eta \geq 0$:

$$\begin{cases} n\partial_\eta \Phi_{\mathbf{J}} = -\langle H_{\mathbf{J}} \rangle_\eta, & \text{(2.7a)} \\ n\partial_\eta^2 \Phi_{\mathbf{J}} = \langle H_{\mathbf{J}} U \rangle_\eta = \langle U^2 \rangle_\eta, & \text{(2.7b)} \\ n\partial_\eta^3 \Phi_{\mathbf{J}} = -\langle U^3 \rangle_\eta, & \text{(2.7c)} \\ n\partial_\eta^4 \Phi_{\mathbf{J}} = \langle U^4 \rangle_\eta - 3\langle U^2 \rangle_\eta^2 - 3 \sum_{i=1}^n \Big\{ \partial_\eta^2 \lambda_i \langle U^2 (x_i - m_i) \rangle_\eta + \dfrac{\partial_\eta^2 \gamma_i}{2} \langle U^2 [x_i^2 - v_i - m_i^2] \rangle_\eta \Big\}. & \text{(2.7d)} \end{cases}$$

In principle one can push this strategy up to arbitrary order $p \geq 1$. Unfortunately, as already noticed in [GY91], we do not know of any closed expression for the order-p derivative. Nevertheless, we are now ready to compute the first orders of the high-temperature (or small-η) expansion of the free entropy. Note that in this expansion the Lagrange parameters $\{\lambda_i(\eta), \gamma_i(\eta)\}$ will always be considered at $\eta = 0$.

First orders of the expansion – First of all, taking $\eta = 0$ one easily reaches from eq. (2.5):

$$\Phi_{\mathbf{J}}(\eta = 0) = \frac{1}{2} \ln 2\pi + \frac{1}{n} \sum_{i=1}^n \Big[\frac{\gamma_i}{2}(v_i + m_i^2) - \frac{1}{2} \ln \gamma_i + \lambda_i m_i + \frac{\lambda_i^2}{2\gamma_i} \Big].$$

After extremization over the Lagrange multipliers $\{\lambda_i, \gamma_i\}$ this yields:

$$\Phi_{\mathbf{J}}(\eta = 0) = \frac{1 + \ln 2\pi}{2} + \frac{1}{2n} \sum_{i=1}^n \ln v_i.$$

At order 1 we have from eq. (2.7):

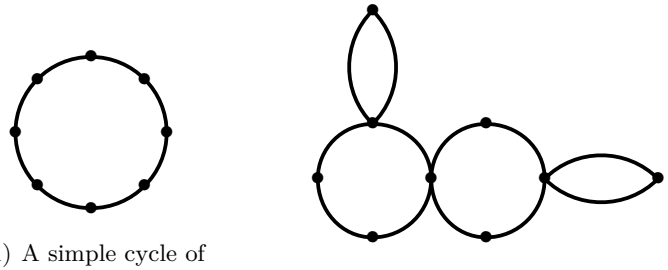
$$\Big(\partial_\eta \Phi_{\mathbf{J}} \Big)_{\eta=0} = -\frac{1}{n} \langle H_{\mathbf{J}} \rangle_{\eta=0} = \frac{1}{2n} \sum_{i,j} J_{ij} m_i m_j + \frac{1}{2n} \sum_{i=1}^n J_{ii} v_i. \tag{2.8}$$

For any η, we can write the Maxwell-type relations:

$$\begin{cases} \gamma_i(\eta) &= n\, \partial_{v_i} \Phi_{\mathbf{J}}(\eta), \\ m_i \gamma_i(\eta) + \lambda_i(\eta) &= n\, \partial_{m_i} \Phi_{\mathbf{J}}(\eta). \end{cases} \tag{2.9}$$

These relations plugged in eq. (2.8) lead to $\partial_\eta \gamma_i(\eta = 0) = J_{ii}$ and $\partial_\eta \lambda_i(\eta = 0) = \sum_{j(\neq i)} J_{ij} m_j$. We then obtain the U operator at $\eta = 0$ from eq. (2.6):

$$U(\eta = 0, \mathbf{J}) = -\frac{1}{2} \sum_{i \neq j} J_{ij}(x_i - m_i)(x_j - m_j). \tag{2.10}$$



(A) A simple cycle of order $p = 8$.

(B) A diagram at order $p = 12$.

FIGURE 2.1: Diagrammatic representation of the perturbative terms. Each vertex represents an index i_α and carries a factor v_{i_α}, while each edge is a factor J_{ij}. Each connected component of the diagram carries a global factor n^{-1}.

We can then directly apply eq. (2.7) to reach the order 2:

$$\frac{1}{2}\left(\partial_\eta^2 \Phi_{\mathbf{J}}\right)_{\eta=0} = \frac{1}{2n}\langle U^2\rangle_{\eta=0} = \frac{1}{4n}\sum_{i\neq j} J_{ij}^2 v_i v_j. \tag{2.11}$$

At order 3 we obtain:

$$\frac{1}{3!}\left(\partial_\eta^3 \Phi_{\mathbf{J}}\right)_{\eta=0} = -\frac{1}{6n}\langle U^3\rangle_{\eta=0} = \frac{1}{6n}\sum_{i,j,k} J_{ij}J_{jk}J_{ki}v_i v_j v_k + \mathcal{O}_n(1),$$

in which the sum is made over *pairwise distinct* i, j, k indices. At order 4 we reach[3]

$$\frac{1}{4!}\left(\partial_\eta^4 \Phi_{\mathbf{J}}\right)_{\eta=0} = \frac{1}{8n}\sum_{i,j,k,l} J_{ij}J_{jk}J_{kl}J_{li}v_i v_j v_k v_l + \mathcal{O}_n(1),$$

where again, i, j, k, l are pairwise distinct indices.

Larger orders – As we mentioned, the PGY expansion can not give analytic results for arbitrary perturbation orders. Nevertheless, our results up to order 4 lead to the natural conjecture:

$$\begin{aligned}\Phi_{\mathbf{J}}(\eta) = \frac{1+\ln 2\pi}{2} + \frac{1}{2n}\sum_{i=1}^{n}\ln v_i + \frac{\eta}{2n}\sum_{i\neq j} J_{ij}m_i m_j \\ + \frac{1}{n}\sum_{p=1}^{\infty}\frac{\eta^p}{2p}\sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distinct}}} J_{i_1 i_2}J_{i_2 i_3}\cdots J_{i_{p-1}i_p}J_{i_p i_1}\prod_{\alpha=1}^{p} v_{i_\alpha}.\end{aligned} \tag{2.12}$$

Note that in order to obtain this formula, we took the $n \to \infty$ limit at every perturbation order in η, which is part of the implicit assumptions of the PGY expansion. The terms of this perturbative expansion can be represented diagrammatically as *simple cycles* of order p, see Fig. 2.1a.

Elementary diagrammatics – Generically, at any order in the expansion one can construct a diagrammatic representation of the contributing terms, as shown in Fig. 2.1. While we will extensively discuss these diagrammatics in Section 2.4, let us note that the only remaining terms in eq. (2.12) correspond to *simple cycles*, as depicted in Fig. 2.1a. For the case of orthogonal couplings, this dominance of simple cycles was already noted in [PP95]. Note that many other diagrams may not be negligible in the limit $n \to \infty$ (e.g. the "cactus" diagram of Fig. 2.1b), however as we will detail in Section 2.4, they will cancel out in the free entropy calculation[4].

[3] For pedagogical purposes we detail this calculation in Appendix A.1.

[4] At order 4, this was shown explicitly in Appendix A.1.

Comparison with the exact solution – In the high temperature phase, the solution to the maximization of eq. (2.12) under $\mathbf{m}$ is the paramagnetic solution $\mathbf{m} = 0$. Furthermore, we expect that the $\{v_i\}$ that maximize the free entropy of eq. (2.12) are *homogeneous*, that is $\forall i,\ v_i = v$. The spherical constraint of eq. (2.4) thus gives $v = \sigma^2$. We can compare the result of the resummation of simple cycles, eq. (2.12) with the exact results of eq. (2.3) in the paramagnetic phase. For these two results to agree, we need the generating function for simple cycles to be related to the $\mathcal{R}$-transform of ρ by:

$$\mathbb{E}\left[\frac{1}{n}\sum_{p=1}^{\infty}\frac{\eta^p\sigma^{2p}}{2p}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distinct}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1}\right] = \frac{1}{2}\int_0^{\eta\sigma^2}\mathcal{R}_\rho(x)\mathrm{d}x + \mathcal{O}_n(1), \tag{2.13}$$

in which the outer expectation is with respect to the distribution of $\mathbf{J}$. In particular, an order-by-order comparison yields that the *free cumulants* $\{c_p(\rho)\}_{p\in\mathbb{N}^\star}$, defined in Section 1.5, satisfy:

$$\forall p \in \mathbb{N}^\star, \quad c_p(\rho) = \lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distinct}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1}\right]. \tag{2.14}$$

Using rigorous results of [GM05], we can prove a stronger version of eq. (2.14), namely convergence in L^2 norm, that we give as a theorem:

Theorem 2.2 (*Simple cycles and free cumulants*)

For any matrix $\mathbf{J} \in \mathcal{S}_n$ generated by Model S and any $p \in \mathbb{N}^\star$, one has:

$$\lim_{n\to\infty}\mathbb{E}\left|\frac{1}{n}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distinct}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1} - c_p(\rho)\right|^2 = 0.$$

We postpone the proof of this result, along with a much more detailed analysis of the diagrammatics, to Section 2.4. Assuming that we can invert the summation over p and the $n\to\infty$ limit, Theorem 2.2 implies we have a stronger version eq. (2.13) with L^2 convergence, for small enough η:

$$\frac{1}{n}\sum_{p=1}^{\infty}\frac{\eta^p\sigma^{2p}}{2p}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distinct}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1} \xrightarrow[n\to\infty]{L^2} \frac{1}{2}\int_0^{\eta\sigma^2}\mathcal{R}_\rho(x)\mathrm{d}x. \tag{2.15}$$

This identity is a "resummation" of the single-graph free entropy of eq. (2.12)! As a final note we can use eq. (2.3) to write the resummation in an alternative form (dropping $\mathcal{O}_n(1)$ terms):

$$\frac{1}{n}\sum_{p=1}^{\infty}\frac{\eta^p\sigma^{2p}}{2p}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distinct}}} J_{i_1i_2}\cdots J_{i_pi_1} = \frac{1}{2}\inf_{\gamma\geq\eta\lambda_{\max}}\left[\gamma\sigma^2 - \int\rho(\mathrm{d}\lambda)\ln(\gamma-\eta\lambda)\right] - \frac{1+\ln\sigma^2}{2}. \tag{2.16}$$

Note that this is true a priori only in the high temperature phase $\eta < \eta_c$, so that both sides are analytic functions of η. In the next paragraph, we will investigate this discontinuity and show that it coincides with the instability of the paramagnetic solution in the TAP free entropy.

Stability of the paramagnetic phase

We can check the stability of the paramagnetic solution for $\eta \leq \eta_c \equiv -\sigma^{-2}\mathcal{S}_\rho(\lambda_{\max})$. Recall that we must satisfy the norm constraint $v = \sigma^2 - (1/n)\sum_i m_i^2$. Solely as a function of $\mathbf{m}$ the free entropy therefore reads, up to $\mathcal{O}_n(1)$ terms:

$$\Phi_{\mathbf{J}}(\eta, \mathbf{m}) = \frac{1+\ln 2\pi}{2} + \frac{1}{2}\ln\Big[\sigma^2 - \frac{1}{n}\sum_{i=1}^n m_i^2\Big] + \frac{\eta}{2n}\sum_{i\neq j} J_{ij}m_i m_j + G_\rho\Big(\eta\Big[\sigma^2 - \frac{1}{n}\sum_{i=1}^n m_i^2\Big]\Big),$$

in which G_ρ is the integrated $\mathcal{R}$-transform of ρ, defined for all $0 \leq x \leq -\mathcal{S}_\rho(\lambda_{\max})$:

$$G_\rho(x) \equiv \frac{1}{2}\int_0^x \mathrm{d}u\, \mathcal{R}_\rho(u). \tag{2.17}$$

The Hessian of the *extensive* free entropy $n\Phi_{\mathbf{J}}$ at the paramagnetic solution $\mathbf{m} = 0$ is thus:

$$\mathrm{Hess}_n(\eta) \equiv n\left(\frac{\partial^2\Phi_{\mathbf{J}}}{\partial m_i \partial m_j}\right)_{\mathbf{m}=0} = -\frac{\delta_{ij}}{\sigma^2}[1 + \eta\sigma^2\mathcal{R}_\rho(\eta\sigma^2)] + \eta J_{ij} + \mathcal{O}_n(1).$$

The paramagnetic solution is stable as long as the Hessian is negative. This is true as long as $\eta < \eta_c$, at which point the spectrum of $\mathrm{Hess}_n(\eta_c)$ touches zero by definition of η_c. Our PGY expansion allows thus to compute the free entropy in the high temperature phase (which is paramagnetic), coherently with the direct computation results. More generically, as shown by Plefka [Ple82] in the SK model, $\mathrm{Hess}_n(\eta)$ is related to the inverse susceptibility matrix of the system, and thus the singularity of the Hessian implies a singularity of the free entropy (i.e. a phase transition).

Validity of the expansion and stability of the replica symmetric solution – Beyond the paramagnetic regime, it is an open question to relate the range of validity η_c of the Plefka expansion and the de Almeida-Thouless condition that characterizes the local stability of the replica symmetric solution (see [DAT78] for its original derivation, and [Kab08a, SK08] for examples of its applications in inference problems). The equivalence of these two conditions was shown in the seminal paper of Plefka [Ple82] in the Sherrington-Kirpatrick model. It is tedious but straightforward to generalize this conclusion to a model with Ising spins $x_i = \pm 1$ and a Hamiltonian given by eq. (2.1) with a rotationally-invariant coupling matrix $\mathbf{J}$. However, investigating the relation between these two conditions in general models appears to be an open problem, and does not enter the scope of this thesis.

The free cumulant series – We performed an expansion of $\Phi_{\mathbf{J}}(\eta)$ close to $\eta = 0$, which implies that this expansion is thus valid in the region $(0, \eta_c)$, in which $\eta_c = -\sigma^{-2}\mathcal{S}_\rho(\lambda_{\max})$ is the first non-analyticity of $\Phi_{\mathbf{J}}(\eta)$, see eq. (2.3) and the discussion above. However there exists spectrums for which the function $\mathcal{R}_\rho(x)$ can be analytically extended beyond $x_c \equiv -\mathcal{S}_\rho(\lambda_{\max})$, e.g. Wigner's semicircle for which $\mathcal{R}_{\text{s.c.}}(x) = x$ and $x_c = 1$. Yet, one has to be careful that this does not imply that the free entropy $\Phi_{\mathbf{J}}(\eta)$ is analytic beyond η_c, and even in this case our Plefka expansion is *a priori* only valid up to $\eta = \eta_c$.

2.2.2 Generalization to a bipartite model

In this section $n, m \geq 1$, $\sigma_x, \sigma_h > 0$, and we will consider the *thermodynamic* limit in which $n, m \to \infty$ with $m/n \to \alpha > 0$. We define the following bipartite Hamiltonian:

$$H_{\mathbf{L}}(\mathbf{h}, \mathbf{x}) \equiv -\mathbf{h}^\intercal \mathbf{L}\mathbf{x} = -\sum_{\mu=1}^m\sum_{i=1}^n L_{\mu i}h_\mu x_i, \qquad \mathbf{h} \in \mathbb{S}^{m-1}(\sigma_h\sqrt{m}), \quad \mathbf{x} \in \mathbb{S}^{n-1}(\sigma_x\sqrt{n}). \tag{2.18}$$

The coupling matrix $\mathbf{L} \in \mathbb{R}^{m\times n}$ is assumed to be rotationally-invariant in the sense of Model R.

Direct free entropy computation

The calculation for this bipartite case is very similar to the direct calculation of Section 2.2.1, and we introduced in Section 1.5.3 the corresponding "rectangular" HCIZ integrals, cf. in particular Theorem 1.13. We obtain that for all values of η the free entropy can be expressed as:

$$\begin{aligned}\Phi_{\mathbf{L}}(\eta) &\equiv \lim_{n\to\infty} \frac{1}{n} \ln \int \mu_m(\mathrm{d}\mathbf{h}) \int \mu_n(\mathrm{d}\mathbf{x})\, e^{\eta \mathbf{h}^\intercal \mathbf{L}\mathbf{x}}, \\ &= \frac{1+\alpha}{2}\ln 2\pi + \frac{1}{2} \inf_{\substack{\gamma_h,\gamma_x>0 \\ \gamma_h\gamma_x \geq \eta^2\lambda_{\max}}} \Big[\alpha\gamma_h\sigma_h^2 + \gamma_x\sigma_x^2 - (\alpha-1)\ln\gamma_h - \int \rho(\mathrm{d}\lambda)\ln(\gamma_x\gamma_h - \eta^2\lambda)\Big],\end{aligned} \tag{2.19}$$

where ρ is the asymptotic eigenvalue distribution of $\mathbf{L}^\intercal\mathbf{L}$ (see Model R).

Plefka-Georges-Yedidia expansion

The PGY expansion for this model is very similar to the symmetric model we just studied. We constraint the first and second moments as $\langle h_\mu\rangle = m_\mu^h$, $\langle x_i\rangle = m_i^x$, $\langle h_\mu^2\rangle = v_\mu^h + (m_\mu^h)^2$ and $\langle x_i^2\rangle = v_i^x + (m_i^x)^2$. The spherical constraints impose $\sigma_h^2 = (1/m)\sum_{\mu=1}^m [v_\mu^h + (m_\mu^h)^2]$ and $\sigma_x^2 = (1/n)\sum_{i=1}^n [v_i^x + (m_i^x)^2]$. As in Section 2.2.1 one can study the diagrams that appear in the Plefka expansion. We show again the L^2 concentration of the simple cycles, and the negligibility of all other diagrams in the expansion. We state in more details these results for the bipartite case in Section 2.4.5. We obtain the following result, a counterpart to eq. (2.12) for bipartite models:

$$\begin{aligned}\Phi_{\mathbf{L}}(\eta) &= \frac{1+\alpha}{2}[1+\ln 2\pi] + \frac{\alpha}{2m}\sum_{\mu=1}^m \ln v_\mu^h + \frac{1}{2n}\sum_{i=1}^n \ln v_i^x + \frac{\eta}{n}\sum_{\mu=1}^m\sum_{i=1}^n L_{\mu i} m_\mu^h m_i^x \\ &+ \frac{1}{n}\sum_{p=1}^\infty \frac{\eta^{2p}}{2p} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distinct}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distinct}}} L_{\mu_1 i_1} L_{\mu_1 i_2} L_{\mu_2 i_2} \cdots L_{\mu_p i_p} L_{\mu_p i_1} \prod_{\alpha=1}^p v_{\mu_\alpha}^h v_{i_\alpha}^x + \mathcal{O}_n(1),\end{aligned} \tag{2.20}$$

in which indices $\{\mu_l\}$ run from 1 to m and indices $\{i_l\}$ run from 1 to n. Assuming uniform variances at the maximum: $v_\mu^h = v^h, v_i^x = v^x$, and comparing to eq. (2.19) in the paramagnetic phase, we obtain the correspondence, in the high temperature phase:

$$\begin{aligned}&\frac{\alpha\ln\sigma_h^2 + \ln\sigma_x^2}{2} + \frac{1}{n}\sum_{p=1}^\infty \frac{\eta^{2p}\sigma_h^{2p}\sigma_x^{2p}}{2p} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distinct}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distinct}}} L_{\mu_1 i_1} L_{\mu_1 i_2} L_{\mu_2 i_2} \cdots L_{\mu_p i_p} L_{\mu_p i_1} \\ &= -\frac{1+\alpha}{2} + \frac{1}{2}\inf_{\gamma_h,\gamma_x} \Big[\alpha\gamma_h\sigma_h^2 + \gamma_x\sigma_x^2 - (\alpha-1)\ln\gamma_h - \int \rho(\mathrm{d}\lambda)\ln(\gamma_h\gamma_x - \eta^2\lambda)\Big].\end{aligned} \tag{2.21}$$

Again, the $n\to\infty$ limit is implicit, and the equality then holds in the sense of L^2 convergence.

2.3 PGY expansion for inference models

In this section we perform Plefka (or PGY) expansions for generic models of pairwise interactions, both symmetric and bipartite, and discuss the iterations of the resulting fixed point equations. First, let us recall the important different approximation schemes studied in Section 1.4.3, namely *Expectation-Consistency* (EC), *adaptive TAP* (adaTAP) and *Vector Approximate Message Passing* (VAMP). These three schemes (that we showed to be equivalent) allow to compute

the free entropy of the inference models we will consider, and will be crucial in our discussion. Indeed, one goal of this chapter is to provide a precise justification for the exactness of these schemes by leveraging random matrix theory. Let us briefly describe our strategy:

- In Section 2.3.1 we generalize the PGY expansion of Section 2.2 to inference models and highlight the main differences and assumptions of our method. This yields a precise and systematic justification of the TAP equations for rotationally invariant models. We apply these results to retrieve the TAP free entropy of the Hopfield model, compressed sensing, as well as Generalized Linear Models (GLMs), flagship high-dimensional inference models that we defined in Section 1.1.
- One then needs to maximize this free entropy, which yields fixed point equations: the celebrated *TAP equations*. Iterating them is in itself a challenge since different choices for the iteration scheme can lead to drastically different convergence properties. In Section 2.3.2 we relate message-passing algorithms, which have proven to be very successful both numerically and for theoretical studies, to the approximation schemes mentioned above, and to the PGY expansion. More precisely, we show that the stationary limit of the message-passing equations yields the TAP equations obtained by the PGY expansion.

2.3.1 PGY expansion in generic models of pairwise interactions

Models of symmetric pairwise interactions

A symmetric model with generic priors – Let us first consider a slight generalization of the model of eq. (2.1): the vector $\mathbf{x}$ is now drawn with independent components, each x_i with distribution P_i. They interact via a pairwise interaction (with a rotationally-invariant coupling matrix $\mathbf{J}$), and are subject to an external field $\mathbf{h}$. At a given inverse temperature $\eta \geq 0$ and a fixed realization of $\mathbf{J}$, the Gibbs-Boltzmann distribution of the spins is given as:

$$P_{\eta,\mathbf{J}}(\mathrm{d}\mathbf{x}) \equiv \frac{1}{Z_{\mathbf{J}}(\eta)} \prod_{i=1}^{n} P_i(\mathrm{d}x_i) \, \exp\Big\{\frac{\eta}{2}\sum_{i,j} J_{ij}x_i x_j - \eta \sum_i h_i x_i\Big\}. \tag{2.22}$$

As before, we compute the large n limit of the free entropy $\Phi_{\mathbf{J}}(\eta) \equiv n^{-1} \ln Z_{\mathbf{J}}(\eta)$ at fixed values of the magnetizations $m_i = \langle x_i \rangle$ and variances $v_i = \langle (x_i - m_i)^2 \rangle$, using a PGY expansion. We impose these constraints with Lagrange multipliers $\{\lambda_i\}$ and $\{\gamma_i\}$. Clearly the zero-th order term in η is different from the spherical case, and is given by:

$$\Phi_{\mathbf{J}}(0) = \frac{1}{n}\sum_i \lambda_i m_i + \frac{1}{2n}\sum_i \gamma_i(v_i + m_i^2) + \frac{1}{n} \ln \int \prod_i P_i(\mathrm{d}x_i) \, e^{-\frac{1}{2}\sum_i \gamma_i x_i^2 - \sum_i \lambda_i x_i}. \tag{2.23}$$

At order 1 in η we obtain at leading order:

$$\Big(\frac{\partial \Phi_J}{\partial \eta}\Big)_{\eta=0} = \frac{1}{2n}\sum_{i,j} J_{ij} m_i m_j + \frac{1}{2n}\sum_i J_{ii} v_i - \frac{1}{n}\sum_i h_i m_i.$$

The operator U of Georges-Yedidia, defined in eq. (2.6), is the same as in the spherical case, that is eq. (2.10). This implies that many of the results obtained for the spherical case will transfer here directly. For instance the second order term is identical and given in eq. (2.11). At third order we obtain:

$$\frac{1}{3!}\Big(\frac{\partial^3 \Phi_J}{\partial \eta^3}\Big)_{\eta=0} = \frac{1}{6n} \sum_{\substack{i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_1} v_{i_1} v_{i_2} v_{i_3} + \frac{1}{6n}\sum_{i \neq j} J_{ij}^3 \kappa_i^{(3)} \kappa_j^{(3)}. \tag{2.24}$$

In this equation, we denoted $\kappa_i^{(p)}$ the *cumulant* of order p of x_i at $\eta = 0$. By the rotation invariance of Model S, the term $\sum_{i\neq j} J_{ij}^3$ gives a negligible contribution to the free entropy. We shall therefore assume that the second part of the RHS of eq. (2.24) is negligible as $n \to \infty$. This is correct provided that the possible correlations of the third order cumulants $\kappa_i^{(3)}$ with $\mathbf{J}$ do not drastically change the scaling of this term to make it thermodynamically relevant. This will be the case at all orders, as argued in Section 2.4: the cumulants of order $p \geq 3$ will not contribute to the asymptotic free entropy! The first term of eq. (2.24) corresponds to a simple cycle of order 3 and is the same term that appeared in the spherical case.

We carry on the computation of the derivatives $\partial_\eta^p \Phi_{\mathbf{J}}(\eta = 0)$. Given our remark above on high-order cumulants, we conjecture that the higher-order terms are different from the spherical model only in terms which are sub-leading in n. We will give more detail on the precise hypothesis behind this conjecture in Section 2.4.4. In the end, we reach the following conjecture for $\Phi_{\mathbf{J}}$:

$$\Phi_{\mathbf{J}}(\eta) = \Phi_{\mathbf{J}}(0) + \frac{\eta}{2n}\sum_{i,j} J_{ij}m_i m_j - \frac{\eta}{n}\sum_i h_i m_i + \frac{1}{n}\sum_{p=1}^{\infty}\frac{\eta^p}{2p}\sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} J_{i_1 i_2}\cdots J_{i_p i_1}\prod_{\alpha=1}^{p} v_{i_\alpha}. \quad (2.25)$$

For the remainder of this section we assume that the maximum of the free entropy of eq. (2.25) is attained for variables $\{v_i\}$ such that $v_i = v$. This hypothesis can be argued as reasonable for many models, and we postpone this argumentation to the specific models we will consider. We obtain a resummation of the Plefka free entropy using the correspondence of eq. (2.15):

$$\Phi_{\mathbf{J}}(\eta) = \Phi_{\mathbf{J}}(0) + \frac{\eta}{2n}\sum_{i,j} J_{ij}m_i m_j - \frac{\eta}{n}\sum_i h_i m_i + \frac{1}{2}\int_0^{\eta v} \mathcal{R}_\rho(u)\,\mathrm{d}u. \quad (2.26)$$

Recall that $\Phi_J(0)$ is given in eq. (2.23). As discussed in the spherical case, we expect this expansion of the free entropy to hold for $\eta < \eta_c$, in which $\eta_c \equiv -v^{-1}\,\mathcal{S}_\rho(\lambda_{\max})$.

The PGY expansion to justify EC approximations – The result of the PGY expansion in eq. (2.25) provides a systematic way to show the exactness of the EC approximations (cf. Section 1.4.3) for rotationally invariant models. For instance, as we saw in eq. (1.57), adaTAP amounts to assuming that at every order $p \geq 1$ of perturbation in η, one can perform the calculation *as if* the statistics of the variables were Gaussian. This statement, which generalizes the Parisi-Potters result [PP95], is exactly what we argued to obtain eq. (2.25), using the diagrammatic analysis of Section 2.4. As such, our analysis provides a clear meaning to the EC approximations, by detailing which diagrams are negligible in the $n \to \infty$ limit. This also justifies that the EC approximations are actually exact asymptotically for rotationally invariant models in the high temperature phase, which we summarized in Conjecture 2.1. We believe that this asymptotic exactness extends beyond the high temperature phase to any model in the replica symmetric phase, and the diagrammatic analysis provides a route to proving this statement.

Application to the Hopfield model – As a first application of our framework, we consider the Hopfield model [Hop82]. We let binary spins $\mathbf{x} \in \{\pm 1\}^n$ and the coupling matrix $\mathbf{J}$ is constructed out of p *patterns*, which are spin configurations $\xi^l \in \{\pm 1\}^n$, for $l \in \{1, \cdots, p\}$:

$$\begin{cases} J_{ij} &= \frac{1}{n}\sum_{l=1}^{p} \xi_i^l \xi_j^l \qquad (i \neq j), \\ J_{ii} &= 0. \end{cases} \quad (2.27)$$

We assume that the $\{\xi_i^l\}$ are i.i.d. variables with equal probability in $\{\pm 1\}$, so that $\mathbb{E}J_{ij} = 0$ and $\mathbb{E}J_{ij}^2 = p/n^2$. We study this system in the limit in which both $p, n \to \infty$ with a fixed ratio $p/n \to \alpha$. The derivation of the TAP free energy for these models has been performed in

[NT97, Méz17] via the Plefka expansion, and via the cavity method in [MPV87]. If the random matrix ensemble of eq. (2.27) is *a priori* not rotationally invariant, one can show that since the variables $\{\xi_i^l\}$ are i.i.d., only the first and second moment of their distributions will contribute to the thermodynamic limit of the free entropy, so that we can assume that they are standard centered Gaussian variables without changing the free entropy. The ensemble of eq. (2.27) is thus for our purposes essentially a Wishart matrix model in which the diagonal has been removed. From Section 1.5, we know that its $\mathcal{R}$-transform reads:

$$\mathcal{R}_{\mathbf{J}}(x) = \frac{\alpha}{1-x} - \alpha = \frac{\alpha x}{1-x}. \tag{2.28}$$

Note that because $x_i \in \pm 1$ the variance is given by $v = 1 - n^{-1}\sum_i m_i^2 = 1 - q$, with q the spin glass order parameter. From eq. (2.26) and eq. (2.28), we reach:

$$\Phi_{\mathbf{J}}(\eta) = -\frac{1}{n}\sum_i \Big[\frac{1+m_i}{2}\ln\frac{1+m_i}{2} + \frac{1-m_i}{2}\ln\frac{1-m_i}{2}\Big] + \frac{\eta}{2}\sum_{i,j} J_{ij}m_i m_j$$
$$- \frac{\alpha\eta(1-q)}{2} - \frac{\alpha}{2}\ln[1-\eta(1-q)] + \mathcal{O}_n(1).$$

Maximizing it over the magnetizations $\{m_i\}$ yields the TAP equations for the Hopfield model:

$$\eta^{-1}\tanh^{-1}(m_i) = \sum_{j(\neq i)} J_{ij}m_j - \alpha\eta\frac{1-q}{1-\eta(1-q)}m_i.$$

This is in agreement with the findings of [MPV87, NT97, Méz17]. However, our framework and results allowed us to treat this kind of model in a very fast and generic way.

Models of bipartite pairwise interactions

A bipartite model with generic priors – As in the symmetric setting, we extend the bipartite model of eq. (2.18) by considering a generic prior on the variables rather than a spherical constraint. More precisely, the fields $\mathbf{h}$ and $\mathbf{x}$ are assumed to follow prior distributions $P_X(\mathrm{d}\mathbf{x}) = \prod_i P_i(\mathrm{d}x_i)$ and $P_H(\mathrm{d}\mathbf{h}) = \prod_\mu P_\mu(\mathrm{d}x_\mu)$. For a fixed $\eta \geq 0$ and a realization of the coupling matrix $\mathbf{L}$ we define the Gibbs-Boltzmann distribution:

$$P_{\eta,\mathbf{L}}(\mathbf{h},\mathbf{x}) \equiv \frac{1}{Z_{\eta,\mathbf{L}}}\prod_{\mu=1}^m P_\mu(\mathrm{d}h_\mu)\prod_{i=1}^n P_i(\mathrm{d}x_i)\,\exp\Big\{\eta\sum_{\mu,i} L_{\mu i}h_\mu x_i\Big\}.$$

As in the symmetric case we compute the large n limit of the free entropy $\Phi_{\mathbf{L}}(\eta) \equiv n^{-1}\ln Z_{\eta,\mathbf{L}}$. Recall that we take the limit $n, m \to \infty$ with $m/n \to \alpha > 0$. We constraint the first and second moments of $\{x_i\}$ and $\{h_\mu\}$ under the Gibbs measure to be $\langle x_i\rangle = m_i^x$, $\langle h_\mu\rangle = m_\mu^h$, $\langle (x_i - m_i^x)^2\rangle = v_i^x$, $\langle (h_\mu - m_\mu^h)^2\rangle = v_\mu^h$. The Lagrange multipliers introduced to enforce these conditions are $\lambda_i^x, \lambda_\mu^h, \gamma_\mu^x, \gamma_\mu^h$. At order 0, we obtain easily:

$$\Phi_{\mathbf{L}}(\eta=0) = \frac{1}{n}\sum_{i=1}^n\Big[\lambda_i^x m_i^x + \frac{1}{2}\gamma_i^x(v_i^x + (m_i^x)^2)\Big] + \frac{1}{n}\sum_{\mu=1}^m\Big[\lambda_\mu^h m_\mu^h + \frac{1}{2}\gamma_\mu^h(v_\mu^h + (m_\mu^h)^2)\Big]$$
$$+ \frac{1}{n}\ln\int\prod_\mu P_\mu(\mathrm{d}h_\mu)\prod_i P_i(\mathrm{d}x_i)\,e^{-\sum_i[\lambda_i^x x_i + \frac{1}{2}\gamma_i^x x_i^2] - \sum_\mu[\lambda_\mu^h h_\mu + \frac{1}{2}\gamma_\mu^h h_\mu^2]}. \tag{2.29}$$

The calculations at the first orders are very similar to the ones we already performed. We obtain:

$$\begin{cases} \Big(\dfrac{\partial \Phi_{\mathbf{L}}}{\partial \eta}\Big)_{\eta=0} &= \dfrac{1}{n}\displaystyle\sum_{\mu,i} L_{\mu i} m_\mu^h m_i^x, \\ \dfrac{1}{2}\Big(\dfrac{\partial^2 \Phi_{\mathbf{L}}}{\partial \eta^2}\Big)_{\eta=0} &= \dfrac{1}{2n}\displaystyle\sum_{\mu,i} L_{\mu i}^2 v_\mu^h v_i^x, \\ \dfrac{1}{3!}\Big(\dfrac{\partial^3 \Phi_{\mathbf{L}}}{\partial \eta^3}\Big)_{\eta=0} &= \dfrac{1}{6n}\displaystyle\sum_{\mu,i} L_{\mu i}^3 \kappa_\mu^{(3,h)} \kappa_i^{(3,x)}, \\ \dfrac{1}{4!}\Big(\dfrac{\partial^4 \Phi_{\mathbf{L}}}{\partial \eta^4}\Big)_{\eta=0} &= \dfrac{1}{8n}\displaystyle\sum_{\mu_1\neq\mu_2}\sum_{i_1\neq i_2} L_{\mu_1 i_1} L_{\mu_1 i_2} L_{\mu_2 i_2} L_{\mu_2 i_1} v_{\mu_1}^h v_{\mu_2}^h v_{i_1}^x v_{i_2}^x + \mathcal{O}_n(1). \end{cases} \tag{2.30}$$

Recall that $\kappa_i^{(p,x)}, \kappa_\mu^{(p,h)}$ are the p-th order cumulant of x_i, h_μ at $\eta = 0$. The discussion on the higher orders in perturbation is very similar to the one we just made in the symmetric case: we only retain the simple cycles made of matrix elements $\{L_{\mu i}\}$. We will detail more the extension of the diagrammatics to the bipartite case and the handling of the higher-order cumulants in Section 2.4.5. We obtain the free entropy at leading order in n:

$$\Phi_{\mathbf{L}}(\eta) = \Phi_{\mathbf{L}}(0) + \frac{\eta}{n}\sum_{\mu,i} L_{\mu i} m_\mu^h m_i^x + \frac{1}{n}\sum_{p=1}^{\infty}\frac{\eta^{2p}}{2p}\sum_{\substack{\mu_1,\cdots,\mu_p \\ i_1,\cdots,i_p}} L_{\mu_1 i_1} L_{\mu_1 i_2}\cdots L_{\mu_p i_p} L_{\mu_p i_1}\prod_{\alpha=1}^{p} v_{\mu_\alpha}^h v_{i_\alpha}^x. \tag{2.31}$$

In the summation above all indices $\mu_1, \cdots, \mu_p$ are pairwise distinct, and so are $i_1, \cdots, i_p$. As in the symmetric case, we now assume that the maximum of the free entropy of eq. (2.31) is attained for variables $\{v_\mu^h, v_i^x\}$ such that $v_\mu^h = v^h$ and $v_i^x = v^x$. Using eq. (2.21), the free entropy can be resummed:

$$\begin{aligned} \Phi_{\mathbf{L}}(\eta) =& \Phi_{\mathbf{L}}(0) + \frac{\eta}{n}\sum_{\mu,i} L_{\mu i} m_\mu^h m_i^x - \frac{1+\ln v^x}{2} - \alpha\Big\{\frac{1+\ln v^h}{2}\Big\} \\ &+ \frac{1}{2}\inf_{\zeta^x,\zeta^h}\Big[\alpha\zeta^h v^h + \zeta^x v^x - (\alpha-1)\ln\zeta^h - \int \rho(\mathrm{d}\lambda)\ln(\zeta^x\zeta^h - \eta^2\lambda)\Big], \end{aligned} \tag{2.32}$$

where ρ is the LSD of $\mathbf{L}^\intercal\mathbf{L}$, and recall that $\Phi_{\mathbf{L}}(0)$ is given by eq. (2.29). As in the symmetric case, the diagrammatic study (performed in Section 2.4.5) plays a decisive role in our analysis.

Compressed Sensing: a first inference problem – Compressed sensing [Don06] is a textbook inference problem with numerous applications, which can be seen as a particular case of GLMs. It can be formulated as the inference of the vector $\mathbf{X}$, generated from a prior P_X, from:

$$Y_\mu = \sum_{i=1}^{n} F_{\mu i} X_i + \sqrt{\Delta} z_\mu. \tag{2.33}$$

In this equation we modeled the noise by $z_\mu \sim \mathcal{N}(0,1)$, and the noise value is $\Delta > 0$. We follow [KMS+12], in which the authors considered i.i.d. matrices $\mathbf{F}$. Defining $\mathbf{J} = -\mathbf{F}^\intercal\mathbf{F}$ and using Bayes' rule we can write the posterior distribution of $\mathbf{x}$ as:

$$P_{\mathbf{F}}(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z_{\mathbf{Y},\mathbf{F}}} P_X(\mathbf{x}) \exp\Big\{\frac{1}{2\Delta}\sum_{i,j} J_{ij}\, x_i x_j + \frac{1}{\Delta}\sum_{\mu,i} F_{\mu i} Y_\mu x_i\Big\}.$$

Defining $\eta \equiv \Delta^{-1}$ this distribution can be matched to the one of eq. (2.22) with $h_i \equiv -\sum_\mu F_{\mu i} Y_\mu$. Using this correspondence we can use eq. (2.32) to directly obtain:

$$\Phi_{\mathbf{Y},\mathbf{F}} = \Phi_{\mathbf{Y},\mathbf{F}}(\Delta = \infty) - \frac{1}{2\Delta n}\sum_{i,j}(\mathbf{F}^\intercal\mathbf{F})_{ij} m_i m_j + \frac{1}{n\Delta}\sum_{\mu,i} F_{\mu i} Y_\mu m_i + \frac{1}{2}\int_0^{v\Delta^{-1}} \mathcal{R}_{\mathbf{J}}(u)\mathrm{d}u. \quad (2.34)$$

We postpone the analysis of the corresponding fixed point equations to our algorithmic discussion in Section 2.3.2.

Generalized Linear Models with correlated matrices – GLMs are of primary importance in a very wide variety of scientific and engineering fields, as we underlined in Section 1.1. Consider $m, n \geq 1$ both going to infinity with a ratio $m/n \to \alpha > 0$. We are given a left and right rotationally invariant measurement matrix $\mathbf{F} \in \mathbb{R}^{m\times n}$, as defined in Model R. Given $\mathbf{F}$, data samples $\{Y_\mu\}$ are generated as:

$$\forall \mu \in \{1, \cdots, m\}, \qquad Y_\mu \sim P_{\text{out}}(\,\cdot\,|(\mathbf{F}\mathbf{X})_\mu), \quad (2.35)$$

in which $\mathbf{X} \in \mathbb{R}^n$ is the vector we try to recover (drawn with i.i.d. coordinates from a prior P_X), and P_{out} is a fixed probabilistic channel. Compressed sensing (2.33) corresponds to a Gaussian channel distribution with zero mean and variance Δ. Recall that we consider the Bayes-optimal setting: P_{out} and P_X are *known*, so we can use them in the posterior distribution

$$P(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z(\mathbf{Y},\mathbf{F})}\prod_{i=1}^n P_X(x_i)\prod_{\mu=1}^m P_{\text{out}}[Y_\mu|(\mathbf{F}\mathbf{x})_\mu].$$

While in compressed sensing $\eta = \Delta^{-1}$ played naturally the role of an inverse temperature, in the general setting of eq. (2.35) there is *a priori* no way to perform a Plefka expansion. As it turns out, there is a way to introduce an auxiliary parameter in terms of which we will perform the expansion, similarly to what is done in [AFP16, Alt18]. Introducing the usual Lagrange parameters to fix the means and variances of $\{x_i\}$, we obtain the free entropy:

$$\Phi_{\mathbf{Y},\mathbf{F}} \equiv \frac{1}{n}\sum_i \lambda_i m_i + \frac{1}{2N}\sum_i \gamma_i(v_i + m_i^2) + \frac{1}{n}\ln\int_{\mathbb{R}^n}\mathrm{d}\mathbf{x}\, e^{-S[\mathbf{x}]},$$

in which we introduced an *action* $S[\mathbf{x}]$:

$$S[\mathbf{x}] \equiv \sum_i \lambda_i x_i + \frac{1}{2}\sum_i \gamma_i x_i^2 - \sum_\mu \ln P_{\text{out}}\Big(Y_\mu\Big|\sum_i F_{\mu i} x_i\Big) - \sum_i \ln P_X(x_i).$$

Letting $\mathbf{h} \equiv \mathbf{F}\mathbf{x}$ and using the Fourier transform of the Dirac distribution we reach:

$$\Phi_{\mathbf{Y},\mathbf{F}} = -\alpha\ln 2\pi + \frac{1}{n}\sum_i \lambda_i m_i + \frac{1}{2n}\sum_i \gamma_i(v_i + m_i^2) + \frac{1}{n}\ln\int \mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{h}\,\mathrm{d}\tilde{\mathbf{h}}\, e^{-S_{\text{eff}}[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}]},$$

with a new effective action S_{eff}:

$$S_{\text{eff}} \equiv \sum_i [\lambda_i x_i + \frac{1}{2}\gamma_i x_i^2 - \ln P_X(x_i)] - \sum_\mu [\ln P_{\text{out}}(Y_\mu|h_\mu) + h_\mu(i\tilde{h}_\mu)] + \sum_{\mu,i}(i\tilde{h}_\mu)F_{\mu i}x_i. \quad (2.36)$$

The key idea is to treat $\mathbf{x}$ and $i\tilde{\mathbf{h}}$ as two independent non-Gaussian fields that interact via the last (quadratic) term of eq. (2.36) and to perform a PGY expansion in terms of this *effective* Hamiltonian, which is exactly the bipartite Hamiltonian of eq. (2.18). We again denote η the

inverse temperature, that is in eq. (2.36) we substitute:

$$\sum_{\mu,i} F_{\mu i}\, x_i\, (i\tilde{h}_\mu) \to \eta \sum_{\mu,i} F_{\mu i}\, x_i\, (i\tilde{h}_\mu),$$

and at the end of the expansion we will set $\eta = 1$. Similarly as for the field $\mathbf{x}$ we will fix the first and second moments of the field $i\tilde{\mathbf{h}}$ as $\langle i\tilde{h}_\mu \rangle_\eta = g_\mu$ and $\langle (i\tilde{h}_\mu)^2 \rangle_\eta = -r_\mu + g_\mu^2$, conditions that will be enforced by new Lagrange parameters $\{\omega_\mu, b_\mu\}$. Although a bit tedious, this is straightforward, and we obtain a free entropy in which we will perform a low-η expansion:

$$\begin{aligned}\Phi_{\mathbf{Y},\mathbf{F}}(\eta) = -\alpha \ln 2\pi + \frac{1}{n}\sum_i \lambda_i m_i + \frac{1}{2n}\sum_i \gamma_i (v_i + m_i^2) + \frac{1}{n}\sum_\mu \omega_\mu g_\mu - \frac{1}{2n}\sum_\mu b_\mu(-r_\mu + g_\mu^2) \\ + \frac{1}{n}\ln\Big\{ \int \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{h}\, \mathrm{d}\tilde{\mathbf{h}}\, e^{-S_{\text{eff}}[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}]} \Big\}.\end{aligned}$$

The effective action and Hamiltonian are expressed as follows:

$$\begin{aligned}S_{\text{eff}}[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}] \equiv \sum_i \lambda_i x_i + \frac{1}{2}\sum_i \gamma_i x_i^2 + \sum_\mu \omega_\mu (i\tilde{h}_\mu) \\ - \frac{1}{2}\sum_\mu b_\mu (i\tilde{h}_\mu)^2 - \sum_i \ln P_X(x_i) - \sum_\mu \ln P_{\text{out}}(Y_\mu|h_\mu) - \sum_\mu h_\mu (i\tilde{h}_\mu) + \eta \overbrace{\sum_{\mu,i} F_{\mu i}\, x_i\, (i\tilde{h}_\mu)}^{H_{\text{eff}}[\mathbf{x},\tilde{\mathbf{h}}]}.\end{aligned} \tag{2.37}$$

From this equation it is clear that the priors on the variables $\{x_i\}$ and $\{(i\tilde{h}_\mu)\}$ decouple. The prior on x_i is $P_X(x_i)$, while the prior distribution on $(i\tilde{h})_\mu$ is related to the Fourier transform of the channel distribution:

$$P_{\tilde{H}}(i\tilde{h}_\mu) \equiv \int \frac{\mathrm{d}h}{2\pi} e^{ih\tilde{h}_\mu} P_{\text{out}}(Y_\mu|h).$$

Using our previous result (2.31) on the PGY expansion for bipartite systems, we conjecture:

$$\begin{aligned}\Phi_{\mathbf{Y},\mathbf{F}}(\eta) = \Phi_{\mathbf{Y},\mathbf{F}}(0) - \frac{\eta}{n}\sum_{\mu i} F_{\mu i} g_\mu m_i \\ + \frac{1}{n}\sum_{p=1}^{\infty} \frac{(-1)^p \eta^{2p}}{2p} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distincts}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} F_{\mu_1 i_1} F_{\mu_2 i_1} \cdots F_{\mu_p i_p} F_{\mu_1 i_p} \prod_{\alpha=1}^{p} r_{\mu_\alpha v_{i_\alpha}} + \mathcal{O}_n(1).\end{aligned} \tag{2.38}$$

We again assume that the maximum of the free entropy of eq. (2.38) will be attained for variance variables $\{v_i, r_\mu\}$ that are *homogeneous*: they satisfy $v_i = v$ and $r_\mu = r$. Using the resummation of eq. (2.32), this leads to a simplified expression:

$$\begin{aligned}\Phi_{\mathbf{Y},\mathbf{F}}(\eta) = \Phi_{\mathbf{Y},\mathbf{F}}(0) - \frac{1}{n}\sum_{\mu i} F_{\mu i} g_\mu m_i - \alpha \frac{1+\ln r}{2} - \frac{1+\ln v}{2} \\ + \frac{1}{2}\inf_{\zeta,\zeta'} \Big\{ \alpha\zeta r + \zeta' v - (\alpha - 1)\ln\zeta - \frac{1}{n}\ln\det[\zeta\zeta' \mathrm{I}_n + \mathbf{F}^\intercal \mathbf{F}] \Big\} + \mathcal{O}_n(1).\end{aligned} \tag{2.39}$$

We will study the fixed point equations corresponding to the free entropy of eq. (2.39) in the following Section 2.3.2.

2.3.2 High-temperature expansions and message-passing algorithms

Several iterations schemes for the TAP equations have been studied in the past literature, and the reader can refer for instance to [OCW16, COFW16]. On a parallel point of view, message-passing algorithms have been extensively studied in the statistical physics literature, as we mentioned in Section 1.4. For i.i.d. matrices it is well understood that the stationary limit of the message-passing iterations is directly related to the fixed point equations of the TAP free entropy. In this section, we extend this correspondence to generic rotationally-invariant matrices and the G-VAMP algorithm [SRF16]. We focus on Generalized Linear Models (GLMs) with matrices $\mathbf{F}$ that satisfy right-rotation invariance, cf. Model R. We first derive the TAP equations from the PGY expansion we performed in Section 2.3.1, before comparing them to the stationary limit of the G-VAMP algorithm.

The TAP equations from the PGY expansion

Following the assumptions of the VAMP and G-VAMP algorithms [SRF16, RSF17] we assume that the variances $\{v_i, r_\mu\}$ are *homogeneous*, that is $r_\mu = r$ and $v_i = v$. We can then use the resummed expression of the Plefka free entropy expressed in eq. (2.39). We first extremize this expression with respect to the Lagrange multipliers $\{\lambda_i, \gamma_i, \omega_\mu, b_\mu\}$:

$$\begin{cases} m_i &= \mathbb{E}_{P_X(\lambda_i,\gamma)}[x], \\ v_i &= \mathbb{E}_{P_X(\lambda_i,\gamma)}[(x-m_i)^2], \\ g_\mu &= -g_{\text{out}}(y_\mu, \omega_\mu, b), \\ r &= -\frac{1}{m}\sum_{\mu=1}^m \partial_\omega g_{\text{out}}(y_\mu, \omega_\mu, b). \end{cases} \tag{2.40}$$

We defined P_X and g_{out} as:

$$\begin{cases} P_X(\lambda_i,\gamma_i)(x) &\propto P_X(x) e^{-\frac{1}{2}\gamma_i x^2 - \lambda_i x}, \\ g_{\text{out}}(y,\omega,b) &\equiv \dfrac{1}{b}\dfrac{\int \mathrm{d}z\, P_{\text{out}}(y|z)\,(z-\omega)\, e^{-\frac{(z-\omega)^2}{2b}}}{\int \mathrm{d}z\, P_{\text{out}}(y|z)\, e^{-\frac{(z-\omega)^2}{2b}}}. \end{cases} \tag{2.41}$$

The remaining equations are obtained by maximizing eq. (2.39) with respect to the physical parameters. We make use of the Jacobi formula for a symmetric positive definite matrix $\mathbf{J}$: $\partial_{J_{ij}}[\ln\det \mathbf{J}] = (\mathbf{J}^{-1})_{ij}$. We reach:

$$\begin{cases} \lambda_i = -\gamma m_i + \sum_\mu F_{\mu i} g_\mu, & \text{(2.42a)} \\ \omega_\mu = g_\mu b + \sum_i F_{\mu i} m_i, & \text{(2.42b)} \\ \zeta \mathcal{S}_{\mathbf{F}^\intercal\mathbf{F}}(-\zeta\zeta') = v, & \text{(2.42c)} \\ \zeta' \mathcal{S}_{\mathbf{F}^\intercal\mathbf{F}}(-\zeta\zeta') = \alpha r - \dfrac{\alpha-1}{\zeta}, & \text{(2.42d)} \\ \gamma = \dfrac{1}{v} - \zeta', & \text{(2.42e)} \\ b = \dfrac{1}{r} - \zeta. & \text{(2.42f)} \end{cases}$$

Additive Gaussian channel – In the case of an additive Gaussian channel with variance Δ we find $r = (\Delta + b)^{-1}$, which gives $\zeta = \Delta$ and $\gamma = \mathcal{R}_{\mathbf{F}^\intercal\mathbf{F}/\Delta}(-v)$. One can check from this that we recover the TAP equations for the compressed sensing problem obtained from eq. (2.34), even though these equations were derived with a "naive" PGY expansion in $\eta \equiv \Delta^{-1}$.

The G-VAMP algorithm

With a similar reasoning that we used to derive the VAMP algorithm for a symmetric pairwise model in Section 1.4.3, we can write a VAMP algorithm for a bipartite model. We do not describe its full derivation here, and we simply report the G-VAMP algorithm for the GLM as stated in [SRF16]. We define a set of functions:

$$\begin{cases} \tilde{F}_m(r,\gamma) \equiv \dfrac{\int \mathrm{d}x P_X(x)\, x\, e^{-\frac{1}{2}\gamma(x-r)^2}}{\int \mathrm{d}x\, P_X(x)\, e^{-\frac{1}{2}\gamma(x-r)^2}}, & \text{(2.43a)} \\ \tilde{F}_v(r,\gamma) \equiv \dfrac{\int \mathrm{d}x P_X(x)\, x^2\, e^{-\frac{1}{2}\gamma(x-r)^2}}{\int \mathrm{d}x\, P_X(x)\, e^{-\frac{1}{2}\gamma(x-r)^2}} - (\tilde{F}_m(r,\gamma))^2, & \text{(2.43b)} \\ \tilde{F}_z(\omega,\tau) \equiv \dfrac{\int \mathrm{d}z P_{\rm out}(y|z)\, z\, e^{-\frac{1}{2}\tau(z-\omega)^2}}{\int \mathrm{d}z P_{\rm out}(y|z)\, e^{-\frac{1}{2}\tau(z-\omega)^2}} = g_{\rm out}(y,\omega,\tau^{-1})\tau^{-1} + \omega, & \text{(2.43c)} \\ \tilde{F}_\kappa(\omega,\tau) \equiv \dfrac{\int \mathrm{d}z P_{\rm out}(y|z)\, z^2\, e^{-\frac{1}{2}\tau(z-\omega)^2}}{\int \mathrm{d}z P_{\rm out}(y|z)\, e^{-\frac{1}{2}\tau(z-\omega)^2}} - (\tilde{F}_z(\omega,\tau))^2 = \partial_\omega g_{\rm out}(y,\omega,\tau^{-1})\tau^{-2} + \tau^{-1}. & \text{(2.43d)} \end{cases}$$

The full algorithm then amounts to iterate the following equations:

$$\mathbf{m}_{1i}^t = \tilde{F}_m((\mathbf{r}_J^t)_i, \gamma_J^t), \qquad v_1^t = \frac{1}{n}\sum_i \tilde{F}_v((\mathbf{r}_J^t)_i, \gamma_J^t), \tag{2.44a}$$

$$\mathbf{r}_0^t = \frac{(\mathbf{m}^t - \gamma_J^t v^t \mathbf{r}_J^t)}{(1-\gamma_J^t v^t)}, \qquad \gamma_0^t = \frac{1}{v^t} - \gamma_J^t, \tag{2.44b}$$

$$\mathbf{z}_{1\mu}^t = \tilde{F}_z((\boldsymbol{\omega}_J^t)_\mu, \tau_J^t), \qquad \kappa_1^t = \frac{1}{n}\sum_\mu \tilde{F}_\kappa((\boldsymbol{\omega}_J^t)_\mu, \tau_J^t), \tag{2.44c}$$

$$\boldsymbol{\omega}_0^t = \frac{(z^t - \tau_J^t \kappa^t \boldsymbol{\omega}_J^t)}{(1-\tau_J^t\kappa^t)}, \qquad \tau_0^{t+1} = \frac{1}{\kappa_1^t} - \tau_J^t, \tag{2.44d}$$

$$\mathbf{m}_2^t = \frac{1}{\tau_0^{t+1}\mathbf{F}^\intercal\mathbf{F} + \gamma_0^t}(\gamma_0^t \mathbf{r}_0^t + \mathbf{F}^\intercal \tau_0^t \boldsymbol{\omega}_0^t), \qquad v_2^t = \frac{1}{n}\mathrm{Tr}\frac{1}{\tau_0^t\mathbf{F}^\intercal\mathbf{F} + \gamma_0^t}, \tag{2.44e}$$

$$\mathbf{r}_J^{t+1} = \frac{(\mathbf{m}_2^t - \gamma_0^t v_2^t \mathbf{r}_0^t)}{(1-\gamma_0^t v_2^t)}, \qquad \gamma_J^{t+1} = \frac{1}{v_2^t} - \gamma_0^t, \tag{2.44f}$$

$$\mathbf{z}_2^t = \mathbf{F}\frac{1}{\tau_0^t\mathbf{F}^\intercal\mathbf{F} + \gamma_0^t}(\gamma_0^t \mathbf{r}_0^t + \mathbf{F}^\intercal \tau_0^t \boldsymbol{\omega}_0^t), \qquad \kappa_2^t = \frac{1}{n}\mathrm{Tr}\{\mathbf{F}^\intercal\mathbf{F}\frac{1}{\tau_0^t\mathbf{F}^\intercal\mathbf{F} + \gamma_0^t}\}, \tag{2.44g}$$

$$\boldsymbol{\omega}_J^{t+1} = \frac{(\mathbf{z}_2^t - \tau_0^t \kappa_2^t \boldsymbol{\omega}_0^t)}{(1-\tau_0^t\kappa_2^t)}, \qquad \tau_J^{t+1} = \frac{1}{\kappa_2^t} - \tau_0^t. \tag{2.44h}$$

TAP equations and fixed points of G-VAMP

We want to see if the stationary limit of G-VAMP, i.e. eq. (2.44) without time indices, is related to the TAP equations we derived with a PGY expansion. At the fixed points of the G-VAMP algorithm, one has in particular the following: $\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}$, $\mathbf{z}_1 = \mathbf{z}_2 = \mathbf{z}$, $v_1 = v_2 = v$ and $\kappa_1 = \kappa_2 = \kappa$. We start from the TAP equations (2.40) and eq. (2.42), and we map them to eq. (2.44):

- From eq. (2.42f) and eq. (2.43d) we can write

$$\frac{1}{b} = \frac{1}{\kappa} - \frac{1}{\zeta}\,, \tag{2.45}$$

  which can be identified with eq. (2.44h), with $b = \tau_J^{-1}$ and $\zeta = \tau_0^{-1}$.

- Using eq. (2.43d) we write eq. (2.42d) as

$$\frac{r}{\zeta'} = -\frac{\kappa}{\zeta' b^2} + \frac{1}{\zeta' b} = \frac{1}{n}\mathrm{Tr}\Big[\frac{1}{\zeta\zeta' + \mathbf{F}^\intercal\mathbf{F}}\Big] + \frac{1-\alpha^{-1}}{\zeta'\zeta}.$$

  Finally from eq. (2.45) we obtain

$$\frac{\kappa}{\zeta} = \frac{1}{\alpha} - \frac{1}{m}\mathrm{Tr}\frac{\zeta\zeta'}{\zeta\zeta' + \mathbf{F}^\intercal\mathbf{F}},$$

  which is compatible with the second part of eq. (2.44g), with $\zeta = \tau_0^{-1}$ and $\zeta' = \gamma_0$.

- Eq. (2.42c) and eq. (2.42e) are equivalent to the second parts of eq. (2.44e) and eq. (2.44f), with $\zeta' = \gamma_0$, $\zeta = \tau_0^{-1}$ and $\gamma = \gamma_J$.

- We write eq. (2.44e) as

$$(\tau_0\mathbf{F}^\intercal\mathbf{F} + \gamma_0)\mathbf{m} = (\gamma_0\mathbf{r}_0 + \mathbf{F}^\intercal\tau_0\boldsymbol{\omega}_0),$$

  and using that $\mathbf{z} = \mathbf{F}\mathbf{m}$, as well as eq. (2.44b) and eq. (2.44d), we arrive at

$$\gamma_J\mathbf{r}_J = \gamma_J\mathbf{m} + \tau_J\mathbf{F}^\intercal(\mathbf{z} - \boldsymbol{\omega}_J),$$

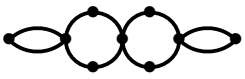
  which is exactly eq. (2.42a) with $\omega = \omega_J$, $\boldsymbol{\lambda} = -\mathbf{r}_J\gamma_J$ and $\tau_J = b^{-1}$.

- Finally we note that eq. (2.42b) at the fixed point is nothing but $\mathbf{z} = \mathbf{F}\mathbf{m}$, which gives eq. (2.44g).

This shows in detail the equivalence between the stationary limit of the G-VAMP algorithm of [SRF16] and the TAP maximization equations that we derived with our PGY expansion!

2.4 Diagrammatics and free cumulants

The goal of this section is to precise how the different diagrams arising in our Plefka expansions can be computed. Recall that for symmetric random matrices $\mathbf{J}$ we construct diagrams as described in Fig. 2.1. For instance the diagram depicted in Fig. 2.1b is:



$$= \frac{1}{n}\sum_{\substack{i_1,\cdots,i_9\\ \text{pairwise distinct}}} (J_{i_1i_2}J_{i_2i_3}J_{i_3i_4}J_{i_4i_1})(J_{i_3i_5}J_{i_5i_6}J_{i_6i_7}J_{i_7i_3})J^2_{i_6i_8}J^2_{i_2i_9}.$$

The perturbation order of any diagram is equal to its number of edges, since each of them represents a factor J_{ij}. The structure of the section is the following:

- In Section 2.4.1 we prove a first rigorous result on the 'simple cycles' arising in the Plefka expansion: we study these diagrams *in expectation over* $\boldsymbol{J}$ and show a weaker version of Theorem 2.2.

- In Section 2.4.2 we extend this study to all possible diagrams, in expectation over $\mathbf{J}$.

- In Section 2.4.3 we show how the results of Section 2.4.1 and Section 2.4.2 can be extended to study the second moments of these diagrams, and use it to show concentration results. This will in particular imply the full statement of Theorem 2.2.

- In Section 2.4.4 we explain how to handle the higher-order moments that can appear as additional factors in these diagrams for the statistical models studied in Section 2.3.

- In Section 2.4.5 we explain how to generalize all these techniques and results to diagrams made of rectangular matrices, that arise in the Plefka expansion for bipartite models.
- Finally, in Section 2.4.6 we show that if one considers an i.i.d. coupling matrix, all the diagrams of order greater than 3 will not contribute in the thermodynamic limit and that one can effectively consider the distribution of the matrix elements to be Gaussian.

Some technicalities, as well as side results and generalizations of these diagrammatics for Hermitian matrices and diverging-size diagrams, which are not directly useful for our expansions, are given in Appendix A.

2.4.1 Expectation of simple cycles and free cumulants

In the following, $\mathbf{J} \in \mathcal{S}_n$ is a *rotationally-invariant* random matrix, cf. Model S. Recall that we defined the *free cumulants* $c_p(\rho)$ of a distribution ρ in Section 1.5. We first show a weaker version of Theorem 2.2:

Theorem 2.3 (*Expectation of simple cycles and free cumulants*)

For any $p \geq 1$, and any set of pairwise distinct indices $i_1, \cdots, i_p \in \mathbb{N}^p$:

$$\lim_{n\to\infty} \mathbb{E}[n^{p-1} J_{i_1 i_2} J_{i_2 i_3} \cdots J_{i_{p-1} i_p} J_{i_p i_1}] = c_p(\rho). \tag{2.46}$$

Actually, we only need to average over $\mathbf{O}$ to obtain the result:

$$n^{p-1} \int_{\mathcal{O}(n)} \mathcal{D}\mathbf{O} \Big[(\mathbf{OJO}^\intercal)_{i_1 i_2} (\mathbf{OJO}^\intercal)_{i_2 i_3} \cdots (\mathbf{OJO}^\intercal)_{i_p i_1} \Big] \xrightarrow[n\to\infty]{\text{a.s.}} c_p(\rho). \tag{2.47}$$

Writing $\mathbf{J} = \mathbf{ODO}^\intercal$ with $\mathbf{D}$ diagonal, the expectation over $\mathbf{O}$ of the *simple loops* considered in the Plefka expansion is an immediate consequence of Theorem 2.3:

$$\forall p \in \mathbb{N}^\star, \quad \mathbb{E}_{\mathbf{O}} \left[\frac{1}{n} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distinct}}} J_{i_1 i_2} J_{i_2 i_3} \cdots J_{i_{p-1} i_p} J_{i_p i_1} \right] \xrightarrow[n\to\infty]{\text{a.s.}} c_p(\rho).$$

Proof of Theorem 2.3 – A first pedestrian way to show eq. (2.47) for small values of p is to use explicit integration of polynomials over the Haar measure of the orthogonal or unitary group. This was first studied by Weingarten [Wei78] and later greatly extended, see for instance [CŚ06]. Performing asymptotic expansions of the resulting so-called Weingarten functions could allow to prove Theorem 2.3, and perhaps even to extend it further, e.g. by a precise description of the rate of the convergence of the simple cycles to the free cumulants. We choose here a different (and arguably simpler) path, leveraging the finite-rank "HCIZ" integrals [HC57, IZ80, GM05] analyzed in Section 1.5.3. Let:

$$L_p^{(n)} \equiv n^{p-1} \int_{\mathcal{O}(n)} \mathcal{D}\mathbf{O} [(\mathbf{OJO}^\intercal)_{i_1 i_2} (\mathbf{ODO}^\intercal)_{i_2 i_3} \cdots (\mathbf{ODO}^\intercal)_{i_p i_1}]. \tag{2.48}$$

To simplify the calculation, we assume that $(i_1, \cdots, i_p) = (1, \cdots, p)$: by rotation invariance $L_p^{(n)}$ does not depend on the particular choice of indices, so this does not remove any generality. The case $p \in \{1, 2\}$ is trivial, so we will assume $p \geq 3$ in the following. One can rewrite eq. (2.48) as:

$$L_p^{(n)} = \frac{1}{n} \prod_{l=1}^{p} \frac{\partial}{\partial b_l} \left[\int_{\mathcal{O}(n)} \mathcal{D}\mathbf{O}\, e^{\frac{n}{2} \mathrm{Tr}\, [\mathbf{M}(\mathbf{b})\mathbf{OJO}^\intercal]} \right]_{\mathbf{b}=0},$$

in which $\mathbf{b} \equiv (b_1, \cdots, b_p)$ and $\mathbf{M}(\mathbf{b}) \in \mathcal{S}_n$ is the following symmetric block matrix of rank p:

$$\mathbf{M}(\mathbf{b}) \equiv \begin{pmatrix} \mathbf{M}_1(\mathbf{b}) & (0) \\ (0) & (0) \end{pmatrix}, \qquad \text{with} \qquad \mathbf{M}_1(\mathbf{b}) \equiv \begin{pmatrix} 0 & b_1 & 0 & \cdots & 0 & b_p \\ b_1 & 0 & b_2 & \cdots & 0 & 0 \\ 0 & b_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & b_{p-1} \\ b_p & 0 & 0 & \cdots & b_{p-1} & 0 \end{pmatrix}.$$

We now apply Theorem 1.12. and we reach (limits are taken a.s. with respect to the law of $\mathbf{J}$):

$$\lim_{n\to\infty} L_p^{(n)} = \lim_{n\to\infty} \frac{1}{n}\Big[\prod_{l=1}^{p} \frac{\partial}{\partial b_l}\Big]\{Z(\mathbf{b})\}_{\mathbf{b}=0}, \quad \text{with} \quad Z(\mathbf{b}) \equiv \exp\left\{\frac{n}{2}\sum_{k=1}^{\infty} \frac{c_k(\rho)}{k} \mathrm{Tr}\,[\mathbf{M}(\mathbf{b})^k]\right\}. \tag{2.49}$$

Note that differentiating $Z(\mathbf{b})$ with respect to b_1 yields by cyclicity of the trace:

$$\frac{1}{Z(\mathbf{b})}\frac{\partial}{\partial b_1} Z(\mathbf{b}) = \frac{n}{2}\sum_{k=1}^{\infty} c_k(\rho) \mathrm{Tr}\,[\mathbf{E}_{12}\mathbf{M}(\mathbf{b})^{k-1}], \tag{2.50}$$

with elementary symmetric matrices $(\mathbf{E}_{ab})_{ll'} \equiv \delta_{l,a}\delta_{l',b} + \delta_{l',a}\delta_{l,b}$. These matrices satisfy $\mathbf{E}_{ab}\mathbf{E}_{cd} = 0$ if $\{c,d\} \cap \{a,b\} = \emptyset$. The only way to obtain a matrix of non-zero trace with a product of matrices $\{\mathbf{E}_{ab}\}$ is thus to have a *cycle* structure in the indices of the matrices. For instance:

$$\mathrm{Tr}\,[\mathbf{E}_{12}^2\mathbf{E}_{13}\mathbf{E}_{23}\mathbf{E}_{12}] = \mathrm{Tr}\,[\mathbf{E}_{12}\mathbf{E}_{21}\mathbf{E}_{13}\mathbf{E}_{32}\mathbf{E}_{21}] \neq 0, \quad \text{while} \quad \mathrm{Tr}\,[\mathbf{E}_{12}^2\mathbf{E}_{24}\mathbf{E}_{23}\mathbf{E}_{12}] = 0.$$

Using this along with $\mathbf{M}(0) = 0$, it is easy to see that the only term that will survive after taking all the successive derivatives and letting $\mathbf{b} = 0$ will be the derivatives of the right-hand-side of eq. (2.50), and not other derivatives of $Z(\mathbf{b})$. Let us analyze what differentiating this term yields. As we saw, differentiating with respect to b_1 yields a matrix $\mathbf{E}_{12}$. When differentiating with respect to b_2 this yields a matrix $\mathbf{E}_{23}$. Note that *a priori*, one has:

$$\frac{\partial}{\partial b_2}\mathrm{Tr}[\mathbf{E}_{12}M(\mathbf{b})^{k-1}] = \mathrm{Tr}\,[\mathbf{E}_{12}\sum_{l=0}^{k-2} M(\mathbf{b})^l \mathbf{E}_{23} M(\mathbf{b})^{k-2-l}]. \tag{2.51}$$

However, the following differentiations with respect to $b_3, \cdots, b_p$ will never yield a matrix of the type $\mathbf{E}_{2a}$. Therefore in eq. (2.51) only two terms, the term $l = 0$ and $l = k-2$, will yield a non-zero contribution. In the end, after taking all the p successive derivatives, only two terms will remain, which correspond to the two possible orientations of the simple cycle:

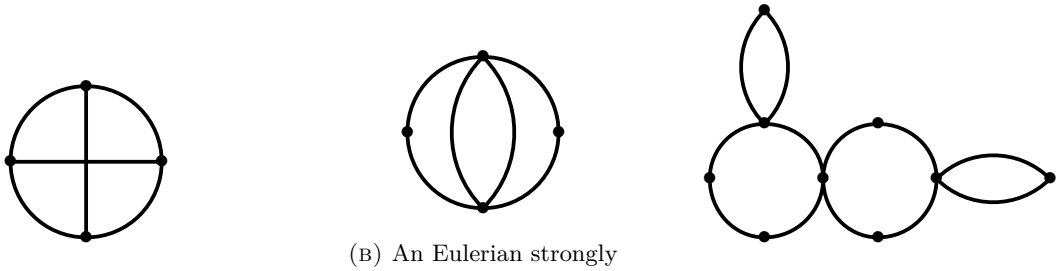
$$\lim_{n\to\infty} L_p^{(n)} = \frac{1}{2}\sum_{k=p}^{\infty} c_k(\rho)\mathrm{Tr}[(\mathbf{E}_{12}\mathbf{E}_{23}\cdots\mathbf{E}_{p1} + \mathbf{E}_{1p}\mathbf{E}_{pp-1}\cdots\mathbf{E}_{32}\mathbf{E}_{21})\ \mathbf{M}(0)^{k-p}] = c_p(\rho),$$

since $\mathbf{M}(0) = 0$. This finishes the proof. □

2.4.2 The expectation of generic diagrams

Following [GY91, PP95] we can define three disjoint categories of connected diagrams:

T.1 *Non-Eulerian diagrams* – By definition, a diagram is Eulerian if one can construct a cyclic path in the graph that goes through each edge exactly once. It is a classic result of graph theory (the Euler–Hierholzer theorem [Eul41, HW73]) that these graphs are exactly the



(A) A non-Eulerian diagram.

(B) An Eulerian strongly irreducible diagram.

(C) A cactus diagram.

FIGURE 2.2: Cactus and non-cactus connected diagrams. Each vertex represents an index i over which we sum, and each edge is a factor J_{ij}. Each diagram carries a global factor n^{-1}.

connected graphs with even degree in each vertex. For instance, the graph depicted in Fig. 2.2a is not Eulerian, whereas the one of Fig. 2.2b is Eulerian.

T.2 *Eulerian diagrams that are strongly irreducible but not simple cycles* – By strongly irreducible [GY91], we mean that one can not make it disconnected by removing any single vertex. E.g. Fig. 2.2b is strongly irreducible, whereas Fig. 2.2c is not.

T.3 *Cactus diagrams* – These diagrams, like the one of Fig. 2.2c, are trees made of simple cycles joining at their vertices. Among them are of course the *simple cycles*.

As generically argued in [GY91], only strongly irreducible diagrams will appear in the PGY expansions. This is an important hypothesis of the Plefka expansion, somehow a bit hidden by the formalism. We give precise descriptions of the large n limit of the *expectation* of all these diagrams in the following[5]. More precisely, we will show:

(*i*) All non-Eulerian diagrams (type T.1) have vanishing expectation in the $n \to \infty$ limit.

(*ii*) All diagrams of type T.2 also have vanishing expectation in the $n \to \infty$ limit.

(*iii*) By Theorem 2.3, the expectation of a simple cycle of size p converges to the p-th free cumulant of ρ. We show that the expectation of a cactus diagram converges to the product of the expectations of all its constituent simple cycles. For instance, for the diagram $\mathcal{C}$ of Fig. 2.2c we obtain that its expectation converges to:

$$\lim_{n\to\infty} \mathbb{E}\,\mathcal{C} = c_2(\rho)^2 c_4(\rho)^2. \tag{2.52}$$

In the remaining of Section 2.4.2, we justify all these claims.

Eulerian diagrams, strongly irreducible diagrams and simple cycles

Let us consider a connected diagram G with V vertices and E edges. We show here that:

- If G is not Eulerian, then $\mathbb{E}\,G \underset{n\to\infty}{\to} 0$
- If G is Eulerian and strongly irreducible but not a simple cycle, then $\mathbb{E}\,G \underset{n\to\infty}{\to} 0$ as well.

As we average over orthogonal matrices the permutation invariance of the indices allows to write:

$$\mathbb{E}\,G = n^{V-1}[1 + o_n(1)] \int_{\mathcal{O}(n)} \mathcal{D}\mathbf{O} \prod_{1\le l<l'\le V} (\mathbf{O}\mathbf{D}\mathbf{O}^\intercal)_{ll'}^{\epsilon_{ll'}},$$

[5] By "expectation" we mean here expectation over $\mathbf{O}$ in $\mathbf{J} = \mathbf{O}\mathbf{D}\mathbf{O}^\intercal$, by rotation invariance.

FIGURE 2.3: Two possible decompositions of the diagram of Fig. 2.2b into simple cycles.

in which the $\epsilon_{ll'}$ are positive integers such that $\sum_{l<l'} \epsilon_{ll'} = E$. We can now use the results of [GM05], as we did in Section 2.4.1, to write this diagram as:

$$\mathbb{E}\, G = n^{V-E-1}[1+\mathcal{O}_n(1)]\Big[\prod_{l<l'} \frac{\partial^{\epsilon_{ll'}}}{\partial b_{ll'}^{\epsilon_{ll'}}}\Big]\left[\exp\left\{\frac{n}{2}\sum_{k=1}^{\infty}\frac{c_k(\rho)}{k}\mathrm{Tr}\,[\mathbf{M}(\mathbf{b})^k]\right\}\right]_{\mathbf{b}=0}. \tag{2.53}$$

Here $M_{ll'}(\mathbf{b}) = M_{l'l}(\mathbf{b}) \equiv b_{ll'}$ for $l < l'$, and $M_{ll}(\mathbf{b}) = 0$. Exactly as in Section 2.4.1, the elementary matrices $\{\mathbf{E}_{ll'}\}$ will appear in eq. (2.53) by successive derivatives of the exponential, using the fact that $\partial_{b_{ll'}}\mathbf{M}(\mathbf{b}) = \mathbf{E}_{ll'}$ and that $\mathbf{M}(0) = 0$. As we explained in Section 2.4.1, a trace of the products of the $\{\mathbf{E}_{ll'}\}$ matrices will only be non-zero if and only if the indices in the products form a cycle. Moreover, as is clear in eq. (2.53), the terms corresponding to the decomposition of $\mathbb{E}\, G$ into the maximum number of such cycles will dominate in the large n limit, as each derivation of the exponential term adds a multiplicative factor n[6]. These two facts together imply that:

- If G is not Eulerian (e.g. Fig. 2.2a) its expectation will be 0 in the limit $n \to \infty$, as by definition it is not possible to decompose it into disjoint cycles.
- If G is Eulerian, strongly irreducible, but not a simple cycle, the dominant contribution to $\mathbb{E}\, G$ in eq. (2.53) will arise from decomposing G into simple cycles, as this decomposition maximizes the number of cycles, and we already showed that each simple cycle has a non-negligible contribution. For the graph of Fig. 2.2b, we show two such possible decompositions in Fig. 2.3.

Given the remarks above we assume now that G is Eulerian and strongly irreducible. Let us denote P the maximal number of simple cycles in such a decomposition of the graph G. Then one can easily see that the scaling of eq. (2.53) will be:

$$\mathbb{E}\, G \sim n^{V+P-E-1}.$$

By elementary graph theory, for a strongly irreducible diagram G we have $V + P - E - 1 \leq 0$, and we have equality iff G is a simple cycle. This implies that all strongly irreducible diagrams that are not simple cycles will not contribute in the $n \to \infty$ limit. This fully justifies claims (i) and (ii) above.

Cactus diagrams

As a side result, although it's not directly useful for our PGY expansion, we show that we can compute the large n limit of any "cactus" [PP95] diagram (like the one of Fig. 2.2c) as a function of the free cumulants of ρ. The argument is straightforward and uses the same technique as in the previous paragraph. Consider a cactus diagram G with V vertices and E edges: one can write exactly eq. (2.53). it is easy to see that there is only *one* maximal decomposition of $\mathbb{E}G$,

[6] There might be a confusion, so we emphasize that this "decomposition" of $\mathbb{E}\, G$ is a decomposition of the *graph* representing $\mathbb{E}\, G$.

which corresponds to its natural decomposition into its constituent simple cycles, and that the number of such cycles is $P = E + V - 1$. Let us denote $\{r_1, \cdots, r_P\}$ the number of vertices in each of these P simple cycles. The dominant contribution corresponds to differentiating P times inside the exponential of eq. (2.53). Using exactly the argument of Section 2.4.1 for each of the P simple cycles we finally obtain the claim (iii) above:

$$\mathbb{E}\, G = n^{P+V-E-1} \prod_{\alpha=1}^{P} c_{r_\alpha}(\rho) + \mathcal{O}_n(1) = \prod_{\alpha=1}^{P} c_{r_\alpha}(\rho) + \mathcal{O}_n(1).$$

2.4.3 Concentration of the diagrams: a second moment analysis

Using the limits of the first moments derived in Sections 2.4.1,2.4.2, we will show:

- (A) If $\mathcal{C}_p$ is the simple cycle of order p, then $\mathcal{C}_p \underset{n\to\infty}{\overset{L^2}{\to}} c_p(\rho)$, which ends the proof of Theorem 2.2. Moreover, if G is a cactus diagram then it converges in the L^2 sense to the products of the free cumulants corresponding to its constituent simple cycles.
- (B) If G is of the type T.1 or T.2 then $\mathbb{E}\, G^2 \underset{n\to\infty}{\to} 0$: G will be negligible in the $n \to \infty$ limit.

As we already mentioned, only strongly irreducible diagrams will appear in the PGY expansion. Together with point (B) this justifies why only the simple cycles contribute in our PGY expansion, as we noticed in Section 2.2.1. In order to show (A) and (B) we will first establish the following fact. Consider a diagram G with V vertices and E edges, of any of the types T.1, T.2, or T.3. Then one has:

$$\mathbb{E}\, G^2 = (\mathbb{E} G)^2 + \frac{1}{n} \sum_{\alpha} \mathbb{E}\, \mathcal{C}_\alpha + \mathcal{O}_n(1). \tag{2.54}$$

In this formula, the sum $\sum_\alpha \mathcal{C}_\alpha$ represents *all the possible diagrams* $\mathcal{C}_\alpha$ that one can obtain by 'gluing' together two replicas of the diagram G. Let us first see why it implies (A) and (B): all diagrams $\mathcal{C}_\alpha$ have a negligible expectation by the first moment analysis we performed in Sections 2.4.1,2.4.2. So for every kind of diagram we considered we have $\mathbb{E}\, G^2 = (\mathbb{E}\, G)^2 + \mathcal{O}_n(1)$. Given our previous computations of the first moments this implies results (A) and (B).
To conclude our argument, we now show eq. (2.54). One can write any diagram G as:

$$G = \frac{1}{n} \sum_{\substack{i_1,\cdots,i_V \\ \text{pairwise distinct}}} \prod_{1\le l<l'\le V} J_{i_l i_{l'}}^{\epsilon_{ll'}},$$

in which the integers $\epsilon_{ll'}$ satisfy $\sum_{l<l'} \epsilon_{ll'} = E$. Thus one has:

$$\mathbb{E}\, G^2 = \mathbb{E}\left[\frac{1}{n^2} \sum_{\substack{i_1,\cdots,i_V \\ \text{pairwise distinct}}} \sum_{\substack{j_1,\cdots,j_V \\ \text{pairwise distinct}}} \prod_{1\le l<l'\le V} J_{i_l i_{l'}}^{\epsilon_{ll'}} J_{j_l j_{l'}}^{\epsilon_{ll'}} \right].$$

In this expression, one can see that two types of terms have to be taken into account:

- Terms for which *all indices* $\{i_1, \cdots, i_V, j_1, \cdots, j_V\}$ are pairwise distinct. Diagrammatically, this corresponds to a graph with two disconnected components that are identical and equal to G. Using the exact same arguments as in Section 2.4.2 and since all the indices are distinct, the decomposition of this diagram into the maximum number of simple cycles will be two copies of the maximal decomposition of G. This shows that the total contribution of these terms behaves asymptotically as $(\mathbb{E}\, G)^2$.

FIGURE 2.4: Second moment decomposition of the simple cycle of order 3. We detail the combinatorial factors.

- Terms for which there is at least one equality of the type $i_l = j_{l'}$ for $1 \leq l, l' \leq V$. Such a term thus corresponds to a diagram with a *single* connected component and constructed by "gluing" some of the vertices of two identical copies of G. Since these diagrams have a single connected component, they carry a single n^{-1} factor, which explains the term $n^{-1} \sum_\alpha \mathbb{E}\, \mathcal{C}_\alpha$ in eq. (2.54), denoting denote $\mathcal{C}_\alpha$ the "glued" diagrams.

This ends our justification of eq. (2.54). We give a schematic representation of this decomposition for G a simple cycle in Fig. 2.4.

2.4.4 Higher-order moments in the diagrammatics

All the diagrammatic results we just derived were valid for diagrams solely made out of the matrix elements $\{J_{ij}\}$, without any additional factors. However in the PGY expansion for non-spherical models there are possible factors that are the cumulants (or the moments) of the variables at $\eta = 0$, as we pointed out in Section 2.3. As an example, let us consider the simple symmetric model of eq. (2.1), but assuming that x_i are independent variables (rather than spherical), with $\kappa_i^{(p)}$ the cumulant of order p of x_i. For instances, two possible contributions to the free entropy at order 6 would be:

$$\begin{cases} n^{-1} \sum\limits_{\substack{i_1,i_2,i_3 \\ \text{pairwise distinct}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1} J_{i_1 i_3}^2 \, v_{i_1}^2 v_{i_2} v_{i_3}^2 v_{i_4}, \\ n^{-1} \sum\limits_{\substack{i_1,i_2,i_3 \\ \text{pairwise distinct}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1} J_{i_1 i_3}^2 \, \kappa_{i_1}^{(4)} v_{i_2} \kappa_{i_3}^{(4)} v_{i_4}. \end{cases} \tag{2.55}$$

Both these contributions are represented by the diagram of Fig. 2.2b! One can see that in order to apply our diagrammatic results to the PGY expansion, we need some additional assumptions:

A.1 By construction of the diagrams, odd cumulants of order greater or equal to 3 only appear in *non-Eulerian* graphs. By the results of Section 2.4.3 we know that such diagrams, without the moments or cumulants as factors, are negligible. We assume that the possible correlations of the cumulants of x_i with the matrix elements $\{J_{ij}\}$ are not strong enough to yield thermodynamically relevant corrections to the free entropy.

A.2 We showed that Eulerian strongly irreducible diagrams that are not simple cycles are negligible. We assume that the higher order moments that appear as additional factors do not change their scaling, so that they remain negligible in the thermodynamic limit.

For instance, A.2 implies that the contributions of both terms in eq. (2.55) are negligible in the $n \to \infty$ limit, as the diagram of Fig. 2.2b is strongly irreducible but is not a simple cycle.

2.4.5 Extension to bipartite models

We detail here how we can treat the diagrams that arise in the Plefka expansion of bipartite models with pairwise interactions. The structure of this section is the following:

- We show first how we can generalize all the techniques and results we developed in the symmetric setting to diagrams constructed from a rotationally-invariant rectangular matrix $\mathbf{L}$ satisfying Model R.

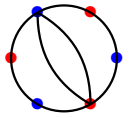


FIGURE 2.5: A diagram constructed from $\mathbf{F}$. Each blue vertex is an index μ, each red vertex an index i. Each edge is a factor $F_{\mu i}$, and each connected component of the diagram carries a global factor n^{-1}. Note that there can only be edges between red and blue vertices.

- We then transpose the assumptions of Section 2.4.4 to this bipartite case, to deal with the higher-order moments of the fields that can arise in the PGY expansion.

Generalization of the previous results to rectangular matrices

Consider a left and right rotationally-invariant random matrix $\mathbf{F} \in \mathbb{R}^{m\times n}$, i.e. satisfying Model R. We are interested in the limit $m, n \to \infty$ with a finite ratio $m/n \to \alpha > 0$. In the PGY expansions performed for bipartite models (e.g. in Section 2.2.2) there appears quantities that we represent as *diagrams*, as explained in Fig. 2.5. The diagram depicted in this figure represents the quantity:

$$\frac{1}{n} \sum_{\substack{\mu_1,\mu_2,\mu_3 \\ \text{pairwise distinct}}} \sum_{\substack{i_1,i_2,i_3 \\ \text{pairwise distinct}}} F_{\mu_1 i_1} F_{\mu_1 i_2} F_{\mu_2 i_2} F_{\mu_2 i_3} F_{\mu_3 i_3} F_{\mu_3 i_1} F_{\mu_1 i_3}^2 .$$

Recall the rectangular transforms and spherical integrals of Section 1.5, in particular the coefficients $\Gamma_p(\alpha, \rho)$ of eq. (1.93). We let $\mathbf{F} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal$, with $\mathbf{U}, \mathbf{V} \in \mathcal{O}(m) \times \mathcal{O}(n)$ two matrices drawn from the Haar measure of their respective group. We can state the counterpart of all our previous results in this rectangular setting:

R.1 Consider a simple cycle of size $2p$. Then it converges in L^2 to $\Gamma_p(\alpha, \rho)$ as $n \to \infty$:

$$\mathbb{E}_{\mathbf{U},\mathbf{V}} \left| \frac{1}{n} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distinct}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distinct}}} F_{\mu_1 i_1} F_{\mu_1 i_2} \cdots F_{\mu_p i_p} F_{\mu_p i_1} - \Gamma_p(\alpha, \rho) \right|^2 \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{2.56}$$

R.2 Any *non-Eulerian* diagram G has vanishing first and second moments: $\lim_{n\to\infty} \mathbb{E}\, G^2 = 0$.

R.3 Any *strongly irreducible* diagram G (i.e. it can not be disconnected by removing a single vertex) that is not a simple cycle (e.g. Fig. 2.5) also has vanishing first and second moments.

R.4 If G is a *cactus*, i.e. a tree made of r simple cycles of sizes $(2p_1, \cdots, 2p_r)$ joining at vertices:

$$\lim_{n\to\infty} \mathbb{E} \left| G - \prod_{l=1}^{r} \Gamma_{p_l}(\alpha, \rho) \right|^2 = 0.$$

Since every argument to show R.1 to R.4 is straightforwardly given by slightly modifying what we did in the symmetric case, we will focus on R.1, and leave the remaining points for the reader.

Justifying R.1 – As in Section 2.4.1, we begin by a first-moment analysis and we show:

$$\mathbb{E} \left[\frac{1}{n} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distinct}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distinct}}} F_{\mu_1 i_1} F_{\mu_1 i_2} F_{\mu_2 i_2} \cdots F_{\mu_p i_p} F_{\mu_p i_1} \right] = \Gamma_p(\alpha, \rho) + \mathcal{O}_n(1). \tag{2.57}$$

Indeed, by rotation invariance we can remove the summations in the LHS of eq. (2.57), and we obtain at leading order in n:

$$\alpha^p n^{2p-1}\mathbb{E}[F_{11}F_{12}F_{22}\cdots F_{pp}F_{p1}] = \frac{1}{n}\frac{\partial^{2p}}{\partial b_1\cdots\partial b_p\partial c_1\cdots\partial c_p}\Big[\int \mathcal{D}\mathbf{U}\mathcal{D}\mathbf{V}e^{\sqrt{\alpha}n\mathrm{Tr}\,[\mathbf{M}(\mathbf{b},\mathbf{c})^\intercal\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal]}\Big]_{\mathbf{b},\mathbf{c}=0},$$

with $\mathbf{M}(\mathbf{b},\mathbf{c})$ a block matrix of rank p defined as:

$$\mathbf{M}(\mathbf{b},\mathbf{c}) = \begin{pmatrix}\mathbf{M}_1(\mathbf{b},\mathbf{c}) & (0)\\ (0) & (0)\end{pmatrix}, \quad \text{with} \quad \mathbf{M}_1(\mathbf{b},\mathbf{c}) \equiv \begin{pmatrix} b_1 & c_1 & 0 & \cdots & 0 & 0\\ 0 & b_2 & c_2 & \cdots & 0 & 0\\ 0 & 0 & b_3 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & b_{p-1} & c_{p-1}\\ c_p & 0 & 0 & \cdots & 0 & b_p\end{pmatrix}.$$

Using Theorem 1.13 we obtain:

$$\alpha^p n^{2p-1}\mathbb{E}[F_{11}F_{12}F_{22}\cdots F_{pp}F_{p1}] = \tag{2.58}$$
$$\frac{1}{n}\frac{\partial^{2p}}{\partial b_1\cdots\partial b_p\partial c_1\cdots\partial c_p}\Big[\exp\Big\{n\sum_{k=1}^{\infty}\frac{\Gamma_k(\alpha,\rho)}{2k}\mathrm{Tr}\,[(\mathbf{M}(\mathbf{b},\mathbf{c})^\intercal\mathbf{M}(\mathbf{b},\mathbf{c}))^k]\Big\}\Big]_{\mathbf{b},\mathbf{c}=0} + \mathcal{O}_n(1),$$

Recall the symmetric elementary matrices $(\mathbf{E}_{ab})_{ij} = \delta_{ai}\delta_{bj} + \delta_{bi}\delta_{aj}$. As in the symmetric case, the dominant terms in eq. (2.58) will be given by the maximum number of differentiations of the exponential term. Moreover, one easily shows $\partial^2_{b_1,c_1}[\mathbf{M}(\mathbf{b},\mathbf{c})^\intercal\mathbf{M}(\mathbf{b},\mathbf{c})] = \mathbf{E}_{12}$. Note that the exponential can only be differentiated once: since $\mathbf{M}(0,0) = 0$, one would need to create cycles with the matrices $\mathbf{E}_{ab}$, and such a cycle can only appear if one derives a single time the exponential term. As in Section 2.4.1, there are two cycles that are created by the successive derivatives: $\mathbf{E}_{12}\mathbf{E}_{23}\cdots\mathbf{E}_{p1}$ and $\mathbf{E}_{21}\mathbf{E}_{1p}\cdots\mathbf{E}_{32}$. These two cycles yield the dominant contribution:

$$\begin{aligned}&\alpha^p n^{2p-1}\mathbb{E}[F_{11}F_{12}F_{22}\cdots F_{pp}F_{p1}]\\ &\quad= \sum_{k=p}^{\infty}\frac{\Gamma_k(\alpha,\rho)}{2}\mathrm{Tr}\,[(\mathbf{E}_{12}\mathbf{E}_{23}\cdots\mathbf{E}_{p1}+\mathbf{E}_{21}\mathbf{E}_{1p}\cdots\mathbf{E}_{32})(\mathbf{M}(0,0)^\intercal\mathbf{M}(0,0))^{k-p}] + \mathcal{O}_n(1),\\ &\quad= \frac{1}{2}\Gamma_p(\alpha,\rho)\mathrm{Tr}\,[\mathbf{E}_{12}\mathbf{E}_{23}\cdots\mathbf{E}_{p1}+\mathbf{E}_{21}\mathbf{E}_{1p}\cdots\mathbf{E}_{32}] + \mathcal{O}_n(1) = \Gamma_p(\alpha,\rho) + \mathcal{O}_n(1).\end{aligned}$$

This shows eq. (2.57). The exact same arguments as the ones used in Section 2.4.3 show that we have L^2 concentration, which concludes our justification of R.1.

The higher order moments and their influence on the diagrammatics

The assumptions we need to make in order to deal with the higher-order moments for rectangular non-spherical models are very similar to A.1 and A.2 (for the symmetric case), and we state them here for completeness:

B.1 From the construction of the diagrams, odd moments of order greater or equal to 3, like $\kappa^{(3,x)}$, only appear in *non-Eulerian* graphs. By R.2 we know that these diagrams (without the moments as factors) are negligible. We assume that the possible correlations of the higher order moments with the matrix elements $\{F_{\mu i}\}$ are not strong enough to yield thermodynamically relevant corrections to the free entropy.

B.2 Eulerian strongly irreducible diagrams that are not simple cycles are negligible by R.3. We assume that the higher order moments that appear as additional factors do not change their scaling, so that they remain negligible in the thermodynamic limit.

2.4.6 A note on i.i.d. matrices

We finish this section by a comment on i.i.d. rectangular matrices. We consider a random matrix $\mathbf{F} \in \mathbb{R}^{m \times n}$ whose elements $\{F_{\mu i}\}$ are taken i.i.d., such that $\sqrt{n}F_{\mu i}$ is drawn from a given probability measure ρ, with zero mean and finite moments of all orders. These matrices appear e.g. in the GAMP algorithm in Section 1.4. If ρ is not a Gaussian probability measure, the matrix $\mathbf{F}$ does not satisfy the rotation invariance of Model R. However, one can still derive strong results on its diagrammatics. Keeping assumptions B.1 and B.2, it is easy to see that because the $\{F_{\mu i}\}$ are uncorrelated, *all diagrams with order $p \geq 3$ are negligible in the $n \to \infty$ limit*. The only diagram that does contribute is:

$$= \frac{1}{n} \sum_{\mu,i} F_{\mu i}^2 v_\mu^h v_i^x. \tag{2.59}$$

In particular, all our PGY expansion formalism remains valid in this case.

Conclusion of Chapter 2

This chapter presented a step-by-step application of the formalism of Georges-Yedidia [GY91] to a wide class of inference problems. It can be seen as a classical example of how "old" methods from statistical mechanics (the TAP formalism and the Plefka expansion were introduced in the 70s and 80s [TAP77, Ple82]) can often provide surprisingly deep and modern results. Our main result is arguably Conjecture 2.1, which we derived in a precise and systematic way by the means of PGY expansions. The diagrammatic results of Section 2.4, as well as the precise forms of the TAP free entropies of Section 2.3, are also of interest, as to the best of our knowledge we did not see in the previous literature such an analysis relating the free cumulants of rotationally-invariant matrices to the diagrams we presented. In the following chapter, we will leverage all the formalism of PGY expansions introduced here to tackle the very involved, and yet open, problem of extensive-rank matrix factorization.

A significant part of the results of [MFC+19] were not developed here, and the interested reader might find there for instance an adaptation of all the techniques of the PGY expansion to study a *replicated system*, i.e. several instances of the system that are correlated through an overlap matrix. This allows to relate the results of the PGY expansion to the ones of the replica method, and to explicit how these two different approaches yield equivalent results. We also describe the connection between AMP equations and the Plefka expansion in the context of generalized linear models with i.i.d. matrices, retrieving the GAMP algorithm [Ran11] and the statistical mechanics analysis of [KMS+12].

Finally, we underline in [MFC+19] a possible important limitation of EC approximations and the PGY expansion with respect to the VAMP approach. Namely, the VAMP approach differs from other EC approximations in the sense that it provides a natural iteration scheme for the equations. This is an important benefit, as iterating these equations is in general a very involved task. For instance, when considering compressive sensing, we showed in Section 2.3 that we could recover the stationary limit of VAMP by a PGY expansion in terms of the inverse noise Δ^{-1}. However, we show in [MFC+19] that there is no clear iteration scheme of the TAP equations that gives back the VAMP algorithm: as physicists have known since forty years, iterating the TAP equations is often itself a hard problem.

Chapter 3

Towards exact solution of extensive-rank matrix factorization

"But there's no sense crying over every mistake,
You just keep on trying till you run out of cake."

GLaDOS, Portal (2007).

Disclaimer – This chapter is devoted to an exciting and involved problem: factorization of extensive-rank matrices. It plays a peculiar role in this thesis, as it is the only chapter solely based on yet unpublished work. For this reason, we will only focus on a very small subset of results, which are on firm enough ground to be presented. While this will leave many open unanswered questions, many of them will be addressed in the upcoming [MFK+21]. We nevertheless provide important results, notably strong arguments disproving several previous approaches to this problem, either based on the replica method [KKM+16] or message-passing algorithms [PSC14a, PSC14b, ZZY20], and we lay out a potential path to correct them.

3.1 Introduction

3.1.1 Definition of extensive-rank matrix factorization

In this chapter we consider the matrix factorization problem in the extensive-rank setting, a problem sometimes referred to as *dictionary learning*. Our approach has similarities with the one of [KKM+16], and an important result of this chapter is a correction to the predictions of this paper. More precisely, we will study the following inference problem:

Model FX (*Extensive-rank matrix factorization*)

Consider $n, m, p \geq 1$. Extensive-rank matrix factorization is defined as the inference of the matrices $\mathbf{F}^\star \in \mathbb{R}^{m\times n}$ and $\mathbf{X}^\star \in \mathbb{R}^{n\times p}$ from the observation of $\mathbf{Y} \in \mathbb{R}^{m\times p}$, generated as:

$$Y_{\mu l} \sim P_{\text{out}}\Big(\,\cdot\,\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} F^\star_{\mu i} X^\star_{il}\Big), \qquad 1\leq \mu \leq m, \quad 1\leq l\leq p.$$

We also assume that the matrix elements of $\mathbf{F}^\star$ and $\mathbf{X}^\star$ are both generated as i.i.d. random variables, with respective prior distributions P_F and P_X, both having zero mean and finite moments of all order.

In order to estimate $\mathbf{F}^\star$ and $\mathbf{X}^\star$ the statistician can use the *posterior distribution*, also referred to as *Gibbs measure* (see Section 1.1 for reminders on these notions)[1]:

$$P_{\mathbf{Y},n}(\mathrm{d}\mathbf{F},\mathrm{d}\mathbf{X}) \equiv \frac{1}{Z_{\mathbf{Y},n}}\prod_{\mu,i} P_F(\mathrm{d}F_{\mu i})\prod_{i,l} P_X(\mathrm{d}X_{il})\prod_{\mu,l} P_{\text{out}}\Big(Y_{\mu l}\Big|\frac{1}{\sqrt{n}}\sum_i F_{\mu i}X_{il}\Big). \tag{3.1}$$

[1] In this chapter, we will generically use greek indices μ,ν for indices between 1 and m, while latin i,j,k indices will run from 1 to n, and the l index between 1 and p.

We assume that she/he has access to the distributions P_{out}, P_F, P_X: this is known as the *Bayes-optimal setting*: it is then known that the mean under the posterior distribution of eq. (3.1) is the information-theoretically optimal estimator.

We consider this inference problem in the high-dimensional (or *thermodynamic*) limit, i.e. we assume $n, m, p \to \infty$ with finite limit ratios $m/n \to \alpha > 0$ and $p/n \to \psi > 0$. In this limit, we can define the *single-graph free entropy* $\Phi_{\mathbf{Y},n}$[2] of the system as:

$$\Phi_{\mathbf{Y},n} \equiv \frac{1}{n(m+p)} \ln \Big[\int \prod_{\mu,i} P_F(\mathrm{d}F_{\mu i}) \prod_{i,l} P_X(\mathrm{d}X_{il}) \prod_{\mu,l} P_{\text{out}}\Big(Y_{\mu l}\Big|\frac{1}{\sqrt{n}}\sum_i F_{\mu i}X_{il}\Big)\Big]. \tag{3.2}$$

The averaged limit free entropy is denoted $\Phi \equiv \lim_{n\to\infty} \mathbb{E}_{\mathbf{Y}}\, \Phi_{\mathbf{Y},n}$. We shall also consider a symmetric version of this problem:

Model $\mathbf{XX}^\intercal$ (*Symmetric extensive-rank matrix factorization*)

Consider $n, m \geq 1$. Symmetric extensive-rank matrix factorization is defined as the inference of the matrix $\mathbf{X}^\star \in \mathbb{R}^{m\times n}$ from the observation of $\mathbf{Y}$ generated as:

$$Y_{\mu\nu} \sim P_{\text{out}}\Big(\,\cdot\,\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X^\star_{\mu i}X^\star_{\nu i}\Big), \qquad 1 \leq \mu < \nu \leq m.$$

We also assume that the elements of $\mathbf{X}^\star$ are generated i.i.d. from a prior distribution P_X.

As in the non-symmetric case, we can define both the posterior distribution of $\mathbf{X}$ and the free entropy of the system as[3]:

$$\begin{cases} P_{\mathbf{Y},n}(\mathrm{d}\mathbf{X}) \equiv \dfrac{1}{Z_{\mathbf{Y},n}} \displaystyle\prod_{\mu,i} P_X(\mathrm{d}X_{\mu i}) \prod_{\mu<\nu} P_{\text{out}}\Big(Y_{\mu\nu}\Big|\frac{1}{\sqrt{n}}\sum_i X_{\mu i}X_{\nu i}\Big), & \text{(3.3a)} \\ \Phi_{\mathbf{Y},n} \equiv \dfrac{1}{nm} \ln \Big[\displaystyle\int \prod_{\mu,i} P_X(\mathrm{d}X_{\mu i}) \prod_{\mu<\nu} P_{\text{out}}\Big(Y_{\mu\nu}\Big|\frac{1}{\sqrt{n}}\sum_i X_{\mu i}X_{\nu i}\Big)\Big]. & \text{(3.3b)} \end{cases}$$

The Gaussian setting – In this chapter, we denote *Gaussian setting* the specific models (both symmetric and non-symmetric) in which all the prior distributions and the channel distributions are zero-mean Gaussians, with variances to be precised.

Possible symmetries of the problem – Assuming that P_x is a zero-mean gaussian distribution, Model **XX**$^\intercal$ has a huge symmetry group: namely one can never distinguish $\mathbf{X}$ from $\mathbf{XO}$, for any $\mathbf{O} \in \mathcal{O}(n)$. A similar symmetry can also arise in Model **FX** provided that both P_F and P_X are zero-mean Gaussians. This symmetry is present as well in the finite-rank case, which was analyzed in [LKZ17]. Note that the zero-mean Gaussian is the *only prior* to satisfy such a property, being the only i.i.d. rotationally invariant distribution in $\mathbb{R}^{m\times n}$. This implies that specifically for this prior, the set of solutions to the TAP equations should possess the same type of symmetry group.

3.1.2 Organization of the chapter and summary of the results

Let us now describe the structure of the rest of Chapter 3.

- We begin in Section 3.2 by a critical analysis of previous approaches to this problem, in the Bayes-optimal setting. In particular, we show that the approach taken in [KKM+16] (as well

[2] We normalize the free entropy by the total number $n(m+p)$ of variables to infer, while the normalization of [KKM+16] is n^2.

[3] Note that one can include the diagonal terms $\mu = \nu$ in these equations without changing the large n limit of the free entropy.

as many subsequent works) is neglecting some crucial correlations of the variables. We provide direct evidence for the influence of these correlations in both the replica calculation and the algorithmic (message-passing) approaches. This implies that adapting these methods (and especially the replica method) for this model is still an open problem.

- In Section 3.3, we tackle the problem using the high-temperature expansion techniques described in Chapter 2. This allows for a systematic computation of the corrections to the results of [KKM+16]. While we restrict here to presenting clear evidence of the existence of such corrections, a final formula for the TAP free entropy of extensive-rank matrix factorization is beyond the scope of this chapter, and will be addressed in [MFK+21].

3.2 Critical treatment of previous approaches

In this section, we briefly sketch the computation of [KKM+16], and detail which approximations used in this work are actually not valid in the thermodynamic limit. The authors derived the free entropy of this problem, in the Bayes-optimal setting, in two different ways, namely via the replica method and via belief propagation (BP) equations. We believe that both their approaches are based on incorrect assumptions.

These wrong hypotheses are also at present in the derivation of the BiGAMP (and BiG-VAMP) algorithm (cf. [PSC14a, PSC14b, ZZY20] among others). We therefore believe that these algorithms are also not able to give exact asymptotic computation of the marginal probabilities in this problem.

Let us now describe both approaches taken in [KKM+16], and explain how the assumptions behind them fail. We focus primarily on the replica analysis, and briefly describe the message-passing approach.

The replica analysis

We focus first on the replica analysis performed in Section V.B of [KKM+16]. As we have introduced in Section 1.3.1, the main idea behind the replica method is to compute the *quenched* free entropy from the evaluation of the moments of the partition function, using the relation:

$$\lim_{n\to\infty}\frac{1}{n^2}\mathbb{E}_{\mathbf{Y}}\ln Z_{\mathbf{Y},n}=\frac{\partial}{\partial r}\Big[\lim_{n\to\infty}\frac{1}{n^2}\ln\mathbb{E}_{\mathbf{Y}}Z^r_{\mathbf{Y},n}\Big]_{r=0}.$$

The computation of the quenched free entropy thus reduces to the evaluation of the integer moments of the partition function. When expanding $\mathbb{E}_{\mathbf{Y}}Z^r_{\mathbf{Y},n}$, there naturally appears $(r+1)$ *replicas* of the system, that interact via the channel distribution term, as represented in the following equation:

$$\mathbb{E}_{\mathbf{Y}}Z^r_{\mathbf{Y},n}=\int\mathrm{d}\mathbf{Y}\Big[\prod_{a=0}^{r}P_F(\mathrm{d}\mathbf{F}^a)P_X(\mathrm{d}\mathbf{X}^a)\Big]\prod_{a=0}^{r}\Big[\prod_{\mu,l}P_{\text{out}}\Big(Y_{\mu l}\Big|\frac{1}{\sqrt{n}}\sum_i F^a_{\mu i}X^a_{il}\Big)\Big].$$

A key step in the calculation of [KKM+16] is the assumption that

$$Z^a_{\mu l}\equiv\frac{1}{\sqrt{n}}\sum_i F^a_{\mu i}X^a_{il} \tag{3.4}$$

are multivariate Gaussian random variables. However, although $\{\mathbf{F}^a\}$ and $\{\mathbf{X}^a\}$ follow statistically independent distributions, with zero mean and finite variances, this is not enough to guarantee the gaussianity of the $Z^a_{\mu l}$ variables in the high-dimensional limit. Indeed, there is a

number $\mathcal{O}(n^2)$ of variables $Z^a_{\mu l}$, and therefore classical central limit results can not conclude on the asymptotic gaussianity of the joint distribution of such variables. In general, this gaussianity is actually false, as can be seen e.g. by considering the following quantity, for a single replica:

$$L_4 \equiv \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n^3}\sum_{\mu_1\neq\mu_2}\sum_{l_1\neq l_2} Z_{\mu_1 l_1} Z_{\mu_1 l_2} Z_{\mu_2 l_2} Z_{\mu_1 l_2}\right].$$

Let us assume that both P_F and P_X are standard Gaussian distributions for the simplicity of the argument. The computation of L_4 is then straightforward and yields $L_4 = \alpha^2\psi^2$. However, should the *joint* distribution of the $\{Z_{\mu l}\}$ converge to a multivariate (zero-mean) Gaussian distribution, such a distribution would satisfy by definition

$$\mathbb{E}[Z_{\mu_1 l_1} Z_{\mu_2 l_2}] = \frac{1}{n}\sum_{i,j} \mathbb{E}[F_{\mu_1 i} F_{\mu_2 j}]\mathbb{E}[X_{l_1 i} X_{l_2 j}] = \delta_{\mu_1\mu_2}\delta_{l_1 l_2}.$$

And using the diagrammatic results of Section 2.4, we know that then $L_4 = 0$ by eq. (2.56) for Gaussian matrices. There is therefore an inconsistency: this argument illustrates the error made (we believe) in [KKM+16] (and many subsequent works) when considering the joint distribution of $\{Z^a_{\mu l}\}$. To conclude on the replica analysis, we did not find a way to correct the calculation, as many of the usual tricks and tools used in the replica method do not transfer here. In particular, the nature of the physical order parameter governing the problem is not clear yet.

The message-passing approach

Another approach to the problem are the Belief Propagation (BP) equations[4], or the *cavity method*. The goal of BP is to compute the *marginal* distributions of each variable in the system, by solving iterative equations involving probability distributions over each single variable. These probability distributions are called *messages* in the BP language, and the fixed point of the iterative equations yields an estimate of the marginal distributions. A detailed treatment of the BP derivation of [KKM+16] is beyond the scope of this chapter, however we believe that the same hidden assumption of Gaussianity is also present in the BP approach, as it neglects the structure of the correlations of some variables.

We will see very precisely in the following Section 3.3 which terms are (wrongly) neglected in the message-passing approach of [KKM+16], and how this is related to neglecting the correlation structure of the variables of eq. (3.4), as in the replica approach.

3.3 TAP equations and PGY expansion

In this section, we detail the Plefka-Georges-Yedidia (PGY) expansion applied to extensive-rank matrix factorization. We mainly focus on detailing the method and its results for the non-symmetric Model **FX**. In Sections 3.3.1 to 3.3.3 we describe the PGY expansion approach to deriving the TAP equations, and we discuss their relation with the previous approaches described in Section 3.2. We end this chapter by Section 3.3.4, in which we generalize our findings to the symmetric Model $\mathbf{XX}^\intercal$.

3.3.1 Sketch of the computation

Rather than the replica method, a more pedestrian way to compute the free entropy of this problem is to perform a perturbative PGY expansion, following the formalism we described in

[4] Recall that we discussed the BP equations on a general factor graph, and their simplifications into approximate message-passing algorithms and their relation with the replica method, in Section 1.4.

Chapter 2. Let us consider the most general Model **FX**. In order to place our problem in a formalism suited for high-temperature expansions, we introduce an auxiliary field $\mathbf{h} \equiv \mathbf{F}\mathbf{X}/\sqrt{n}$ in eq. (3.2), and use the Fourier transform of the resulting delta functions[5]. Introducing delta functions in eq. (3.2), we reach:

$$e^{n(m+p)\Phi_{\mathbf{Y},n}} = \int \prod_{\mu,i} P_F(\mathrm{d}F_{\mu i}) \prod_{i,l} P_X(\mathrm{d}X_{il}) \prod_{\mu,l} \mathrm{d}\hat{H}_{\mu l} P_{\text{out}}(Y_{\mu l}|\hat{H}_{\mu l}) \delta\Big(\hat{H}_{\mu l} - \frac{1}{\sqrt{n}} \sum_i F_{\mu i} X_{il}\Big).$$

Using the Fourier transform of the delta function $\delta(x) = (2\pi)^{-1} \int \mathrm{d}h\, e^{ihx}$, we reach an effective free entropy in terms of three fields $\mathbf{F}, \mathbf{X}, \mathbf{H}$, with $\mathbf{H}$ of size $m \times p$:

$$\Phi_{\mathbf{Y},n} \equiv \frac{1}{n(m+p)} \ln \int \prod_{\mu,l} P_H^{\mu l}(\mathrm{d}H_{\mu l}) \prod_{\mu,i} P_F(\mathrm{d}F_{\mu i}) \prod_{i,l} P_X(\mathrm{d}X_{il}) e^{-H_{\text{eff}}[\mathbf{F},\mathbf{X},\mathbf{H}]}, \tag{3.5}$$

in which we introduced an *effective Hamiltonian* $H_{\text{eff}}[\mathbf{F}, \mathbf{X}, \mathbf{H}]$ and un-normalized probability distributions $\{P_H^{\mu l}\}$ defined as:

$$\begin{aligned} H_{\text{eff}}[\mathbf{F}, \mathbf{X}, \mathbf{H}] &\equiv \frac{1}{\sqrt{n}} \sum_{\mu,i,l} (iH)_{\mu l}\, F_{\mu i}\, X_{il}, \\ P_H^{\mu l}[\mathrm{d}H] &\equiv \int \frac{\mathrm{d}\tilde{H}}{2\pi} e^{iH\tilde{H}} P_{\text{out}}(Y_{\mu l}|\tilde{H}). \end{aligned} \tag{3.6}$$

As we have seen in Chapter 2, the formalism of [GY91] allows to compute the free entropy of the system, constraining the means and variances of each variable $\{F_{\mu i}, X_{il}, H_{\mu l}\}$ by "tilting" the Gibbs measure of eq. (3.5). More formally, we impose:

$$\begin{cases} \langle F_{\mu i} \rangle = m^F_{\mu i}, & \langle X_{il} \rangle = m^X_{il}, \\ \langle F^2_{\mu i} \rangle = v^F_{\mu i} + (m^F_{\mu i})^2, & \langle X^2_{il} \rangle = v^X_{il} + (m^X_{il})^2, \\ \langle (iH)_{\mu l} \rangle = -g_{\mu l}, & \langle (iH)^2_{\mu l} \rangle = -r_{\mu l} + g^2_{\mu l}. \end{cases} \tag{3.7}$$

Recall that $\langle \cdot \rangle$ denotes an average over the (now tilted) Gibbs measure. The resulting free entropy is a function of these means and variances $\{\mathbf{m}^F, \mathbf{m}^X, \mathbf{v}^F, \mathbf{v}^X, \mathbf{g}, \mathbf{r}\}$, on which we will then have to maximize. The conditions of eq. (3.7) will be imposed via Lagrange multipliers, which we denote respectively by $\{\boldsymbol{\lambda}^F, \boldsymbol{\gamma}^F, \boldsymbol{\lambda}^X, \boldsymbol{\gamma}^X, \boldsymbol{\omega}, \mathbf{b}\}$. The free entropy can now be expressed as a function of all these parameters (we still denote it $\Phi_{\mathbf{Y},n}$, with a slight abuse of notations):

$$\begin{aligned} n(m+p)\Phi_{\mathbf{Y},n} = &\sum_{\mu,i} \Big[\lambda^F_{\mu i} m^F_{\mu i} + \frac{\gamma^F_{\mu i}}{2}\Big(v^F_{\mu i} + (m^F_{\mu i})^2\Big)\Big] + \sum_{i,l} \Big[\lambda^X_{il} m^X_{il} + \frac{\gamma^X_{il}}{2}\Big(v^X_{il} + (m^X_{il})^2\Big)\Big] \\ &+ \sum_{\mu,l} \Big[-\omega_{\mu l} g_{\mu l} - \frac{b_{\mu l}}{2}\Big(-r_{\mu l} + g^2_{\mu l}\Big)\Big] + \ln \int P_H(\mathrm{d}\mathbf{H}) P_F(\mathrm{d}\mathbf{F}) P_X(\mathrm{d}\mathbf{X})\, e^{-S_{\text{eff}}[\mathbf{F},\mathbf{X},\mathbf{H}]}, \end{aligned} \tag{3.8}$$

in which we introduced an *effective action* S_{eff}:

$$\begin{aligned} S_{\text{eff}}[\mathbf{F}, \mathbf{X}, \mathbf{H}] \equiv &\sum_{\mu,i} \Big[\lambda^F_{\mu i} F_{\mu i} + \frac{\gamma^F_{\mu i}}{2} F^2_{\mu i}\Big] + \sum_{i,l} \Big[\lambda^X_{il} X_{il} + \frac{\gamma^X_{il}}{2} X^2_{il}\Big] \\ &+ \sum_{\mu,l} \Big[\omega_{\mu l}(iH)_{\mu l} - \frac{b_{\mu l}}{2}(iH)^2_{\mu l}\Big] + H_{\text{eff}}[\mathbf{F}, \mathbf{X}, \mathbf{H}]. \end{aligned} \tag{3.9}$$

[5] Note that we used the same representation for generalized linear models in Section 2.3.

Similarly as what we did in Section 2.3.1 for Generalized Linear Models, the idea of the expansion is to introduce a parameter $\eta > 0$ in front of the interaction Hamiltonian, i.e. to replace $H_{\text{eff}} \to \eta H_{\text{eff}}$ in eq. (3.9), and then take $\eta = 1$ after performing a small-η expansion. We denote the corresponding free entropy $\Phi_{\mathbf{Y},n}(\eta)$. Note that at $\eta = 0$, all the fields $\{F_{\mu i}, X_{il}, H_{\mu\nu}\}$ are independent. In order to make the PGY expansion tractable in this model, we introduce the following hypotheses on the scaling of the parameters:

H.1 At their physical value, the variables $\{m^F_{\mu i}, m^X_{il}\}$ should be uncorrelated, in coherence with the fact that the elements of $\mathbf{F}$ and $\mathbf{X}$ are drawn i.i.d. Importantly, this is not true for $\mathbf{H}$: although the elements $H_{\mu l}$ are independent by eq. (3.6), their distribution depends on μ, l and therefore their statistics might be correlated, e.g. the variables $g_{\mu l}$ might be correlated, through the correlation of the variables $Y_{\mu l}$.

H.2 Recall that $g_{\mu l}$ is the average of $(iH)_{\mu l}$, the conjugate variable to $Z_{\mu l} \equiv \sum_i F_{\mu i} X_{il}/\sqrt{n}$. We therefore assume that the correlations of the elements $\{g_{\mu l}\}$ scale as the ones of $\{Z_{\mu l}\}$, that is *the correlations of the $g = \{g_{\mu l}\}$ scale similarly as the product of two independent matrices that have i.i.d. zero-mean components*[6].

Assuming H.1 and H.2, we derive in Appendix A.3 the following result[7]:

Result 3.1 (*First orders of the PGY expansion for Model FX*)

We have, at leading order as $n \to \infty$:

$$n(m+p)[\Phi_{\mathbf{Y},n}(\eta) - \Phi_{\mathbf{Y},n}(0)] = \frac{\eta}{\sqrt{n}} \sum_{\mu,i,l} g_{\mu l} m^F_{\mu i} m^X_{il} - \frac{\eta^2}{2n} \sum_{\mu,i,l} r_{\mu l}[v^F_{\mu i}(m^X_{il})^2 + (m^F_{\mu i})^2 v^X_{il}]$$
$$+ \frac{\eta^2}{2n} \sum_{\mu,\nu,i} (g^2_{\mu l} - r_{\mu l}) v^F_{\mu i} v^X_{il} + \frac{\eta^4}{4n^2} \sum_{\mu_1,\mu_2} \sum_{l_1,l_2} g_{\mu_1 l_1} g_{\mu_2 l_1} g_{\mu_2 l_2} g_{\mu_1 l_2} v^F_{\mu_1 i} v^F_{\mu_2 i} v^x_{il_1} v^x_{il_2} + \mathcal{O}(\eta^5).$$

Recall moreover that $\Phi_{\mathbf{Y},n}(0)$ is given by

$$n(m+p)\Phi_{\mathbf{Y},n}(0) = \sum_{\mu,i} \Big[\lambda^F_{\mu i} m^F_{\mu i} + \frac{\gamma^F_{\mu i}}{2}\Big(v^F_{\mu i} + (m^F_{\mu i})^2\Big) + \ln \int \mathrm{d}f\, P_F(f)\, e^{-\frac{\gamma^F_{\mu i}}{2} f^2 - \lambda^F_{\mu i} f}\Big] \tag{3.10}$$
$$+ \sum_{i,l} \Big[\lambda^X_{il} m^X_{il} + \frac{\gamma^X_{il}}{2}\Big(v^X_{il} + (m^X_{il})^2\Big) + \ln \int \mathrm{d}x\, P_X(x)\, e^{-\frac{\gamma^X_{il}}{2} x^2 - \lambda^X_{il} x}\Big]$$
$$+ \sum_{\mu,l} \Big[- \omega_{\mu l} g_{\mu l} - \frac{b_{\mu l}}{2}\Big(- r_{\mu l} + g^2_{\mu l}\Big) + \ln \int \mathrm{d}z\, \frac{e^{-\frac{1}{2b_{\mu l}}(z-\omega_{\mu l})^2}}{\sqrt{2\pi b_{\mu l}}} P_{\text{out}}(Y_{\mu l}|z)\Big],$$

in which one has to extremize with respect to all Lagrange multipliers.

It is enlightening to consider the fixed point equations that result from the extremization of the free entropy of Result 3.1. Note that the Lagrange multipliers only appear in the $\eta = 0$ part of the free entropy (this is a general consequence of the formalism of [GY91], that we noticed already in Chapter 2), so we can easily write the maximization over these parameters:

$$\begin{cases} m^F_{\mu i} = \mathbb{E}_{P_F(\lambda^F_{\mu i}, \gamma^F_{\mu i})}[F], & v^F_{\mu i} = \mathbb{E}_{P_F(\lambda^F_{\mu i}, \gamma^F_{\mu i})}[(F - m^F_{\mu i})^2], \\ m^X_{il} = \mathbb{E}_{P_X(\lambda^X_{il}, \gamma^X_{il})}[X], & v^X_{il} = \mathbb{E}_{P_X(\lambda^X_{il}, \gamma^X_{il})}[(X - m^X_{il})^2], \\ g_{\mu l} = g_{\text{out}}(Y_{\mu l}, \omega_{\mu l}, b_{\mu l}), & r_{\mu l} = -\partial_\omega g_{\text{out}}(Y_{\mu l}, \omega_{\mu l}, b_{\mu l}). \end{cases} \tag{3.11}$$

[6] Accounting for such correlations was missing in the previous computations we mentioned in Section 3.2.
[7] We make extensive use of the technicalities of the PGY expansion that we derived in Chapter 2.

We defined $P_F(\lambda,\gamma), P_X(\lambda,\gamma)$ and g_{out} in eq. (2.41) in the previous chapter. We now focus on the equations obtained from the maximization over the "physical" parameters.

3.3.2 The series at order 2 and the approximation of [KKM+16]

We first examine the very first orders of the perturbation series of Result 3.1. E.g. at order 2 in η, the free entropy is:

$$n(m+p)[\Phi_{\mathbf{Y},n}(\eta)-\Phi_{\mathbf{Y},n}(0)] = \frac{\eta}{\sqrt{n}}\sum_{\mu,i,l} g_{\mu l} m^F_{\mu i} m^X_{il} - \frac{\eta^2}{2n}\sum_{\mu i l} r_{\mu l}[v^F_{\mu i}v^X_{il} + v^F_{\mu i}(m^X_{il})^2 + (m^F_{\mu i})^2 v^X_{il}] + \frac{\eta^2}{2n}\sum_{\mu,i,l} g^2_{\mu l} v^F_{\mu i} v^X_{il} + \mathcal{O}(\eta^3). \tag{3.12}$$

Since they are taken at $\eta = 0$, the fixed point equations of eq. (3.11) are unchanged. The maximization over the physical parameters can be done and yields (we indicate on the right of the equation the corresponding parameter over which we maximized):

$$\begin{cases} b_{\mu l} = \frac{\eta^2}{n}\sum_i [v^F_{\mu i}v^X_{il} + v^F_{\mu i}(m^X_{il})^2 + (m^F_{\mu i})^2 v^X_{il}], & (r_{\mu l}) \\ \omega_{\mu l} = \frac{\eta}{\sqrt{n}}\sum_i m^F_{\mu i} m^X_{il} - g_{\mu l}[b_{\mu l} - \frac{\eta^2}{n}\sum_i v^F_{\mu i} v^X_{il}], & (g_{\mu l}) \\ \gamma^F_{\mu i} = \frac{\eta^2}{n}\sum_l [r_{\mu l} v^X_{il} + r_{\mu l}(m^X_{il})^2 - g^2_{\mu l} v^X_{il}], & (v^F_{\mu i}) \\ \lambda^F_{\mu i} = -\frac{\eta}{\sqrt{n}}\sum_l g_{\mu l} m^X_{il} + m^F_{\mu i}[-\gamma^F_{\mu i} + \frac{\eta^2}{n}\sum_l r_{\mu l} v^X_{il}], & (m^F_{\mu i}) \\ \gamma^X_{il} = \frac{\eta^2}{n}\sum_\mu [r_{\mu l} v^F_{\mu i} + r_{\mu l}(m^F_{\mu i})^2 - g^2_{\mu l} v^F_{\mu i}], & (v^X_{il}) \\ \lambda^X_{il} = -\frac{\eta}{\sqrt{n}}\sum_\mu g_{\mu l} m^F_{\mu i} + m^X_{il}[-\gamma^X_{il} + \frac{\eta^2}{n}\sum_\mu r_{\mu l} v^F_{\mu i}]. & (m^X_{il}) \end{cases} \tag{3.13}$$

The reader can check easily that the combined equations (3.11) and (3.13) are actually completely equivalent to the GAMP equations derived in [KKM+16], taking $\eta = 1$ and replacing the notations of the variables following the dictionary:

$$\begin{array}{llll} b_{\mu l} \to V_{\mu l} & \omega_{\mu l} \to \omega_{\mu l} & g_{\mu l} \to g_{\text{out}}(Y_{\mu l}, \omega_{\mu l}, V_{\mu l}) & r_{\mu l} \to -\partial_\omega g_{\text{out}}(Y_{\mu l}, \omega_{\mu l}, V_{\mu l}) \\ \gamma^F_{\mu i} \to Z^{-1}_{\mu i} & \lambda^F_{\mu i} \to -\dfrac{W_{\mu i}}{Z_{\mu i}} & m^F_{\mu i} \to \hat{f}_{\mu i} & v^F_{\mu i} \to s_{\mu i} \\ \gamma^X_{il} \to \Sigma^{-1}_{il} & \lambda^X_{il} \to -\dfrac{T_{il}}{\Sigma_{il}} & m^X_{il} \to \hat{x}_{il} & v^X_{il} \to c_{il} \end{array}$$

In conclusion, our PGY expansion truncated to order η^2 gives back exactly the stationary limit of the GAMP algorithm of [KKM+16]! However, as we will see below, the higher order corrections of order η^4 (and beyond) cannot be neglected: this shows explicitly how the GAMP equations of [KKM+16] (and we believe as well as the BiGAMP equations of e.g. [PSC14a, PSC14b, ZZY20], which are based on the same approximation) are missing important correlations of the problem.

3.3.3 Going to higher orders: open directions

While the order η^2 truncation of Result 3.1 yields back the approximation of [KKM+16], we have computed via our PGY expansions the order η^4, which we recall:

$$\frac{1}{4!n(m+p)}\frac{\partial^4 \mathbf{\Phi}_{\mathbf{Y},n}}{\partial\eta^4}(\eta=0) = \frac{1}{4n^2}\sum_i \sum_{\mu_1,\mu_2}\sum_{l_1,l_2} g_{\mu_1 l_1} g_{\mu_2 l_1} g_{\mu_2 l_2} g_{\mu_1 l_2} v^F_{\mu_1 i} v^F_{\mu_2 i} v^x_{i l_1} v^x_{i l_2} + \mathcal{O}_n(1) \tag{3.14}$$

A crucial observation is that in general, under the natural hypothesis H.2, the term of eq. (3.14) is not negligible as $n \to \infty$! This can be easily seen using our diagrammatic results of Chapter 2,

more precisely eq. (2.56). Indeed, assuming e.g. that the variances are all equal $v^F_{\mu i} = v^F$, $v^X_{il} = v^X$, we have:

$$\frac{1}{4!n(m+p)}\frac{\partial^4 \mathbf{\Phi}_{\mathbf{Y},n}}{\partial \eta^4}(\eta = 0) = \frac{\psi}{\alpha+\psi}\frac{(v^F)^2(v^X)^2}{4}\Gamma_2\Big(\frac{\alpha}{\psi}, \frac{\mathbf{g}^\intercal\mathbf{g}}{n}\Big) + \mathcal{O}_n(1).$$

in which recall that the coefficients Γ_p are functions of the spectrum of the matrix, and were introduced in Section 1.5. They play a role very similar to the free cumulants for symmetric random matrices. Therefore, assuming that the bulk of eigenvalues of $\mathbf{g}^\intercal\mathbf{g}/n$ stays of order 1 as $n \to \infty$ (which is a natural scaling given Hypothesis H.2), the order 4 term gives a non-negligible contribution to the free entropy! This shows in detail how the approximation of [KKM+16] breaks down in this case.

However, our computation does not give all the orders of perturbation, as Result 3.1 is limited to order η^4. As we saw in Chapter 2 this is an intrinsic limitation of the PGY method: there is no generic expression for an arbitrary order of perturbation. However, one can use the first orders to conjecture the form of the higher-order terms. The precise analysis of all these terms is still under investigation, and will be a subject of [MFK+21]. We end this discussion by a few remarks on the results presented here:

- **Similarity with finite-rank problems** – The watchful reader would have noticed that the first orders of Result 3.1 are very similar the the TAP free entropy of GLMs that we derived in Chapter 2, more precisely eq. (2.38). A crucial difference is that here the role of the sensing matrix is played by $-\mathbf{g}$, which is itself a parameter of the TAP free entropy. While we do not yet have an expression for the higher orders in Result 3.1, this similarity is already striking. Importantly, the resummation of eq. (2.39) then suggests that the fully-expanded free entropy might be expressed solely in terms of the *singular value distribution* of $\mathbf{g}/\sqrt{n}$.

- **Nature of the order parameter** – In the Gaussian setting, one can perform exact calculations that leverage extensive-rank HCIZ integrals (cf. Section 1.5.3). While these calculations are still under investigation on our part, and are therefore not presented in this thesis, they suggest that the order parameter governing the state of the model is a *spectral density*. This corroborates our intuition (cf. the previous remark) that the full TAP free entropy might be expressed in terms of the singular value distribution of $\mathbf{g}/\sqrt{n}$. This is an important difference from the finite-rank case: for rank-k matrix factorization (or in general for rank-k recovery problems), the state of the system is governed by a $k \times k$ overlap matrix. For instance, such a matrix order parameter will be crucial to understand simple neural networks in Chapter 4.

- **Iterating the equations** – As we discussed in Section 2.3, the TAP equations are in general not sufficient to obtain an algorithm with good convergence properties. A classical example is given by the Generalized Vector Approximate Passing algorithm (G-VAMP) [SRF16], for which the corresponding TAP equations are derived in Section 2.3.1. As highlighted there, the TAP equations correspond to the stationary limit of G-VAMP, however *there is no obvious iterative resolution scheme of the TAP equations that gives the GVAMP algorithm.* This indicates that even if one obtains all the orders of perturbations in Result 3.1, turning them into an efficient algorithm will pose a serious challenge. We will discuss further these points in [MFK+21].

3.3.4 Symmetric matrix factorization

Performing the Plefka-Georges-Yedidia expansion for Model **XX**$^\intercal$ is extremely similar to the calculation done in previous sections for Model **FX**: one can introduce a field $\mathbf{h} = \mathbf{X}\mathbf{X}^\intercal/\sqrt{n}$, and then performs the same calculations via the Fourier transform of the delta function. In

the following of Section 3.3.4, we give the results of our derivation, while more details are given in Appendix A.4. We adopt similar notations to the ones of Section 3.3.1, removing the X, F subscripts. More precisely, we impose the first and second moment constraints, for $\mu < \nu$ and $i = 1, \cdots, n$:

$$\begin{cases} \langle X_{\mu i} \rangle = m_{\mu i}, & \langle X_{\mu i}^2 \rangle = v_{\mu i} + m_{\mu i}^2, \\ \langle (iH)_{\mu\nu} \rangle = -g_{\mu\nu}, & \langle (iH)_{\mu\nu}^2 \rangle = -r_{\mu\nu} + g_{\mu\nu}^2. \end{cases} \tag{3.15}$$

Note that we will sometimes symmetrize the quantities involved, e.g. we write $g_{\mu\nu} \equiv g_{\nu\mu}$ for $\mu > \nu$, and moreover we adopt the convention $g_{\mu\mu} = r_{\mu\mu} = 0$.

First orders of the PGY expansion

Similarly to Result 3.1, we obtain the first orders of the free entropy as:

Result 3.2 (*First orders of the PGY expansion for Model* $\boldsymbol{XX^\intercal}$)

At leading order as $n, m \to \infty$:

$$\begin{aligned} nm[\Phi_{\mathbf{Y},n}(\eta) - \Phi_{\mathbf{Y},n}(0)] &= \frac{\eta}{\sqrt{n}} \sum_{\substack{\mu<\nu \\ i}} g_{\mu\nu} m_{\mu i} m_{\nu i} - \frac{\eta^2}{2n} \sum_{\substack{\mu<\nu \\ i}} r_{\mu\nu} [v_{\mu i} v_{\nu i} + v_{\mu i} m_{\nu i}^2 + m_{\mu i}^2 v_{\nu i}] \\ &+ \frac{\eta^2}{4n} \sum_{\mu,\nu,i} g_{\mu\nu}^2 v_{\mu i} v_{\nu i} + \frac{\eta^3}{6n^{3/2}} \sum_i \sum_{\substack{\mu_1,\mu_2,\mu_3 \\ \text{pairwise distinct}}} g_{\mu_1\mu_2} g_{\mu_2\mu_3} g_{\mu_3\mu_1} v_{\mu_1 i} v_{\mu_2 i} v_{\mu_3 i} \\ &+ \frac{\eta^4}{8n^2} \sum_i \sum_{\substack{\mu_1,\mu_2,\mu_3,\mu_4 \\ \text{pairwise distinct}}} g_{\mu_1\mu_2} g_{\mu_2\mu_3} g_{\mu_3\mu_4} g_{\mu_4\mu_1} v_{\mu_1 i} v_{\mu_2 i} v_{\mu_3 i} v_{\mu_4 i} + \mathcal{O}(\eta^5). \end{aligned}$$

Recall that the term $\Phi_{\mathbf{Y},n}(0)$ contains the dependency on the channel and priors contributions, as well as the Lagrange multipliers introduced to enforce the conditions of eq. (3.15). Its precise form can be easily deduced from its counterpart in the non-symmetric case, eq. (3.10).

Higher-order terms, and breakdown of previous approximations

One can check that the approximation of [KKM+16, PSC14a, PSC14b, ZZY20] can be adapted easily to Model $\mathbf{XX}^\intercal$, and as in Section 3.2, this approximation amounts to truncating the perturbation series of Result 3.2 at order η^2. However, as in Section 3.3.3, the higher-order terms in Result 3.2 are in general non-negligible. Under a similar hypothesis as H.2, they are actually related to the *free cumulants* of $\mathbf{g}/\sqrt{n}$. This is a consequence of Theorem 2.2, so that assuming e.g. $v_{\mu i} = v$, we have at $\eta = 0$:

$$\begin{cases} \dfrac{1}{3!nm} \partial_\eta^3 \Phi_{\mathbf{Y},n} &= \dfrac{v^3}{6n^{3/2}m} \displaystyle\sum_{\substack{\mu_1,\mu_2,\mu_3 \\ \text{pairwise distinct}}} g_{\mu_1\mu_2} g_{\mu_2\mu_3} g_{\mu_3\mu_1} = \frac{v^3}{6} c_3\Big(\frac{\mathbf{g}}{\sqrt{n}}\Big) + \mathcal{O}_n(1), \\ \dfrac{1}{4!nm} \partial_\eta^4 \Phi_{\mathbf{Y},n} &= \dfrac{v^4}{8n^2m} \displaystyle\sum_{\substack{\mu_1,\mu_2,\mu_3,\mu_4 \\ \text{pairwise distinct}}} g_{\mu_1\mu_2} g_{\mu_2\mu_3} g_{\mu_3\mu_4} g_{\mu_4\mu_1} = \frac{v^4}{8} c_4\Big(\frac{\mathbf{g}}{\sqrt{n}}\Big) + \mathcal{O}_n(1). \end{cases} \tag{3.16}$$

Again, in general these free cumulants are non-negligible in the limit $n \to \infty$: our PGY expansion allowed us to precisely point out the breakdown of previous approximations.

Conclusion of Chapter 3

In this chapter, we presented an application of the very generic formalism of high-temperature PGY expansions (introduced in Chapter 2) to the problem of extensive-rank matrix factorization. Our main finding is that the previous statistical-physics-based approaches to this problem [KKM+16, PSC14a, PSC14b, ZZY20] rely on an approximation that fails in this setting. Via the PGY expansions, we are able to derive systematic corrections to this approximation. While the present chapter does not provide a complete investigation of our corrected equations, these results already raise several important possible generalizations or extensions:

- Of course, the most natural open question is to generalize Results 3.1,3.2 to arbitrary orders of perturbation. Let us focus on Model $\mathbf{X}\mathbf{X}^\intercal$ for the sake of this argument. One could be tempted to generalize eq. (3.16) to any order in perturbation by conjecturing that the order k is related to the k-th free cumulant of $\mathbf{g}/\sqrt{n}$. The resulting expansion is then very similar to the finite-rank series of eq. (2.12), with $\mathbf{g}$ playing the role of the coupling matrix. However, justifying such a wild conjecture (provided it is correct, which is far from obvious) is quite involved, and will be one of the main subjects of [MFK+21].

- In the Gaussian setting, using the results of Matytsin [Mat94] presented in Section 1.5.3, one can perform exact calculations of the free entropy. Although these exact calculations are here very involved, they would be an important check of our PGY expansion. We have already advanced significantly on this line of work, and it will be an important part of [MFK+21].

- It would also be interesting to perform a perturbative calculation (for instance in the symmetric model $\mathbf{X}\mathbf{X}^\intercal$), with a slowly-diverging rank $r = n/m$ (expanding the free entropy in powers of r). Similarly to the PGY expansion, this could provide another systematic correction to the results of [KKM+16], with a different point of view than the one presented in this chapter. Such a calculation is related to an important open problem in random matrix theory, that is a sharp description of the transition between the low-rank spherical HCIZ integral of Theorem 1.14 and the extensive-rank case described in Theorem 1.15. Solving this later problem would help to describe the transition between low-rank and extensive-rank results in the PGY expansion.

- Another interesting open problem is to relate the PGY-expanded free entropy described in Section 3.3 to other approximation schemes. Indeed, as we discussed in Chapter 2, in finite-rank problems these expansions are equivalent to other techniques, e.g. the Expectation Consistency or adaptive TAP approaches. This equivalence does not seem to easily transfer to the extensive-rank case: understanding how to apply these approaches here is another interesting direction of research.

Part II

Physics joins probability: all you need for optimal estimation

A selection of high-dimensional problems

Chapter 4

The physics of learning in a two-layers neural network

"Do you remember the question that caused the creators to attack us, Tali'Zorah? Does this unit have a soul?"

Legion, Mass Effect 3 (2012).

Disclaimer – In this chapter, we begin our tour of high-dimensional problems by a simple model of neural networks. More precisely, we will unveil and rigorously prove phase transitions in the algorithmic and information-theoretic optimal performances in a two-layers neural network known as the *committee machine* in the statistical physics literature. We leverage the replica method, and we put the replica results on rigorous grounds using sophisticated probabilistic interpolation techniques. We will also apply our message-passing toolbox to design an AMP algorithm for this problem, that achieves the optimal performance among a large class of iterative algorithms. All in all, this chapter can be considered as a textbook application of a large part of the statistical physics machinery that we introduced in Chapter 1. It is based on the published work [AMB+19].

4.1 Introduction: the committee machine

While the traditional approach to learning and generalization follows the Vapnik-Chervonenkis [Vap98] and Rademacher [BM02, AAKZ20] worst-case type bounds, there has been a considerable body of theoretical work on calculating the generalization ability of neural networks for data arising from a probabilistic model. Such settings fit well within the framework of statistical mechanics, and the 1990s saw a burst of such studies applied to neural networks [SST92, WRB93, MZ95a, MZ95b, EVdB01]. As the modern success of neural networks has only increased the need to understand their effectiveness [ZBH+16], it is of interest to revisit the results that have emerged thanks to the physics perspective. This direction has been experiencing an important revival in the past years, see e.g. [BJSG+18, BKM+19, CCS+19, Gab20], and this thesis as a whole inscribes itself in this line of work.

As we discussed in Section 1.1, the physics approach is particularly suited to study these models in the so-called *teacher-student* setting. Precisely, labels are generated by feeding i.i.d. random samples to a neural network architecture (the *teacher*) and are then presented to another neural network (the *student*) that is trained using these data. Early studies computed the information theoretic limitations of the supervised learning abilities of the teacher weights by a student who is given m independent n-dimensional examples with $\alpha \equiv m/n = \Theta(1)$ and $n \to \infty$ [SST92, WRB93, EVdB01]. These works relied on non-rigorous heuristic approaches, such as the replica and cavity methods, cf. Sections 1.3.1 and 1.4. Additionally no provably efficient

algorithm was provided to achieve the predicted learning abilities, and it was thus difficult to test those predictions, or to assess the computational difficulty.

Recent developments in statistical estimation and information theory —in particular of message-passing algorithms (cf. [DMM09, Ran11, BM11, JM13] and Section 1.4), and a rigorous proof of the replica formula for the free entropy [BKM+19]— allowed to settle these two missing points for single-layer neural networks (i.e. without any hidden variables).

In the present chapter we leverage many of these works, and provide rigorous asymptotic predictions and corresponding message-passing algorithm for a wide class of two-layers networks. While our results hold for a rather large class of non-linear activation functions, we illustrate our findings on a case considered most commonly in the early literature: the *committee machine*. This is possibly the simplest version of a two-layers neural network where all the weights in the second layer are fixed to unity, and we illustrate it in Fig. 4.1. Denoting Y_μ the label associated with a n-dimensional sample X_μ, and W^*_{il} the weight connecting the i-th coordinate of the input to the l-th node of the hidden layer, it is defined by:

$$Y_\mu = \text{sign}\Big[\sum_{l=1}^{K} \text{sign}\Big(\sum_{i=1}^{n} X_{\mu i} W^*_{il}\Big)\Big] . \tag{4.1}$$

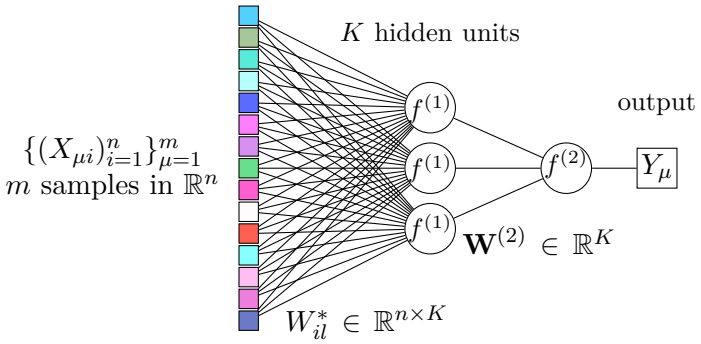


FIGURE 4.1: The *committee machine* is one of the simplest models of two-layers neural networks. Both activation functions are sign functions $f^{(1)}, f^{(2)} = \text{sign}$ and the second-layer weights $\mathbf{W}^{(2)}$ are all fixed to 1. Here we represent a generic two-layers neural network with $K = 3$.

In our teacher-student scenario, the teacher generates i.i.d. data samples $X_{\mu i} \sim \mathcal{N}(0, 1)$, then she/he generates the associated labels Y_μ using a committee machine as in eq. (4.1), with weights W^*_{il} which are unknown to the student. The goal of the student is to learn the weights W^*_{il} from the available examples $(X_\mu, Y_\mu)_{\mu=1}^m$ in order to reach the smallest possible generalization error (i.e. to be able to predict the label the teacher would generate for a new sample not present in the training set).

4.1.1 Classical physics predictions

There have been several studies of this model within the non-rigorous statistical physics approach in the limit where $\alpha \equiv m/n = \Theta(1)$, $K = \Theta(1)$ and $n \to \infty$. In particular, the committee machine attracted a lot of interest during the 1990s, and the reader may refer e.g. to the following series of works: [Sch93, SH92, SH93, MP92, MZ95b, EVdB01]. A particularly interesting result in the teacher-student setting is the *specialization of hidden neurons* (see Sec. 12.6 of [EVdB01], or [SS95] in the context of online learning). These works predicted that for $\alpha < \alpha_{\text{spec}}$ (where α_{spec} is a certain critical value of the sample complexity), the permutational symmetry between hidden neurons remains conserved even after an optimal learning, and the learned weights of

each of the hidden neurons are identical. For $\alpha > \alpha_{\rm spec}$, however, this symmetry gets broken as each of the hidden units correlates strongly with one of the hidden units of the teacher. Another remarkable result obtained in this early literature is the calculation of the optimal generalization error as a function of α.

4.1.2 Main contributions of this chapter

In this chapter we will go beyond the mentioned literature in two main directions. Our first contribution consists in a proof of the replica-symmetric formula for the free entropy, which was conjectured in the statistical physics literature. This proof uses an *adaptive interpolation method* [BM19a, BKM+19], that allows to put several of these results on a rigorous basis. Secondly, we design an Approximate-Message-Passing-type algorithm (recall that we introduced such algorithms in Section 1.4) that is able to achieve the optimal generalization error for a wide range of parameters. The study of AMP —that is widely believed to be optimal among all polynomial-time algorithms in the above setting [DJM13, DM15, ZK16, BPW18]— unveils, in the case of the committee machine with a large number of hidden neurons, the existence a large *hard phase*. In this phase, learning is information-theoretically possible, leading to a generalization error decaying asymptotically as $\mathcal{O}(K/\alpha)$ (in the $\alpha = \Theta(K)$ regime), but where AMP fails and provides only a poor generalization that does not go to zero when increasing α. This strongly suggests that no efficient algorithm exists in this hard region and therefore that there is a computational gap in learning in such neural networks. In other problems where a hard phase was identified its study boosted the development of algorithms that are able to match the predicted thresholds, and we hope that this study will contribute in this regard.

4.2 Main theoretical results

4.2.1 General probabilistic model

Our theoretical analysis is performed for a class of models much broader than the specific committee machine of eq. (4.1). Namely we consider

Model 4.1 (*"Generalized" committee machine, Bayes-optimal setting*)

The observer is given m input samples $\{(X_{\mu i})_{i=1}^n\}_{\mu=1}^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and a probabilistic output channel $P_{\rm out}(y|z)$. Our aim is to recover a set of *teacher weights* $\mathbf{W}^\star \in \mathbb{R}^{n\times K}$, generated i.i.d. from a prior distribution P_0, from the observations

$$Y_\mu \sim P_{\rm out}\Big(\,\cdot\,\Big|\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{\mu i}W^*_{il}\Big\}_{l=1}^K\Big).$$

We assume that P_0 has zero mean, and we denote $\rho \equiv \mathbb{E}[W_0 W_0^\intercal]$ its covariance matrix. As we know, in the Bayes-optimal setting, the best strategy for the student is to learn $\mathbf{W}^*$ from the data $(\mathbf{X}_\mu, Y_\mu)_{\mu=1}^m$ by computing the marginal means of the posterior probability distribution.

A word on notations – In this chapter, we will often manipulate quantities that live either in $\mathbb{R}^K$ or in $\mathbb{R}^{K\times K}$. To lighten the notations, we do not use bold symbols for vectors and matrices of size K, while we keep bold symbols for quantities that have diverging size with n.

Alternative view – Another equivalent way to generate the observations is via

$$Y_\mu = \varphi_{\rm out}\Big(\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{\mu i}W^*_{il}\Big\}_{l=1}^K, A_\mu\Big), \tag{4.2}$$

where $\varphi_{\text{out}} : \mathbb{R}^K \times \mathbb{R} \to \mathbb{R}$ is a generic functions, and $(A_\mu)_{\mu=1}^m$ are i.i.d. real valued random variables with known distribution P_A, that form the probabilistic part of the model, generally accounting for noise.

Different scenarii fit into the general framework of Model 4.1. Among those, the committee machine of eq. (4.1) is obtained when choosing $\varphi_{\text{out}}(h) = \text{sign}(\sum_{l=1}^K \text{sign}(h_l))$. Another renowned model is the parity machine, which corresponds to $\varphi_{\text{out}}(h) = \prod_{l=1}^K \text{sign}(h_l)$, see e.g. [EVdB01]. We will not investigate this last model in this chapter, but the interested reader can refer to [AMB+19] in which we conducted this analysis. A number of layers beyond two has also been considered in specific regimes in [MP92]. Using other activation functions φ_{out}, one can describe many more problems, e.g. compressed pooling [EARK+18, ARK+19] or multi-vector compressed sensing [ZBK16].

4.2.2 Picking from the toolbox

Now that we fixed our model of interest, we can pick our favorite tool from the statistical physics toolbox to study optimal learning. In this chapter, we leave aside the TAP approach (the reader interested in the latter can refer to Chapters 2 and 3) and we focus instead on two other methods.

- First, we leverage the replica method that we introduced in Section 1.3.1. Recall that the aim of this method is the compute the large n limit of the *free entropy* f_n of Model 4.1, i.e. the log-normalization of the posterior that is written in eq. (4.5). It allows to obtain an explicit (conjectural) expression of f_n in the high-dimensional limit $n, m \to \infty$ with $\alpha = m/n$ fixed, called the *replica-symmetric (RS) formula*. One can then naturally ask: can we prove said conjecture? While proving the replica method itself seems out of reach, in this chapter we manage to prove its prediction by use of probabilistic methods: this is summarized by Theorem 4.1, which is the main theoretical result of this chapter. For this reason, while we first derived the results of Theorem 4.1 with the replica method, we focus in this chapter on its probabilistic proof, while the replica computation is detailed in Appendix B.1.

- The second tool, complementary to the first, is the use of *message-passing algorithms* to assess the optimal algorithmic (among a large class of first-order methods, see Section 1.4.2) performance. We introduced these approaches in Section 1.4, and we will extend them to the present setting in Section 4.3.

We hope that this chapter will illustrate to the reader that by leveraging our toolbox one can gain a precise understanding of the computational and statistical optimal performances in a broad class of inference models, and that it can be instrumental in gaining a deeper comprehension of learning in neural networks.

4.2.3 Main theorem: the replica-symmetric formula

Some notation – Recall that $\mathcal{S}_K^+$ is the set of semi-definite positive real symmetric $K \times K$ matrices. For any $M \in \mathcal{S}_K^+$, we can uniquely define its square root $\sqrt{M} = M^{1/2}$. Let us define for $N \in \mathcal{S}_K^+(\mathbb{R})$, the set $S_K^+(N) \equiv \{M \in S_K^+ \text{ s.t. } N - M \in \mathcal{S}_K^+\}$. Note that $\mathcal{S}_K^+(N)$ is both convex and compact.

Two auxiliary inference problems

Stating the RS formula requires introducing two simpler K-dimensional estimation problems:

- The first one is a function of the prior P_0 of Model 4.1. It consists in retrieving a K-dimensional vector $W_0 \sim P_0$ from the K-dimensional observations $Y_0 = r^{1/2} W_0 + Z_0$, with $Z_0 \sim \mathcal{N}(0, \mathrm{I}_K)$

and the "channel gain" matrix $r \in \mathcal{S}_K^+$. The posterior distribution on w is given by

$$\mathbb{P}(w|Y_0) = \frac{1}{\mathcal{Z}_{P_0}} P_0(w) e^{Y_0^\intercal r^{1/2} w - \frac{1}{2} w^\intercal r w}, \tag{4.3}$$

and the associated *free entropy* is given by $\psi_{P_0}(r) \equiv \mathbb{E} \ln \mathcal{Z}_{P_0}$.

- The second problem is a function of the channel P_{out} of Model 4.1. We consider K-dimensional i.i.d. vectors $V, U^* \sim \mathcal{N}(0, \mathrm{I}_K)$ where V is considered to be known and one has to retrieve U^* from the observation of

$$\widetilde{Y}_0 \sim P_{\text{out}}(\,\cdot\,|q^{1/2} V + (\rho - q)^{1/2} U^*),$$

with $q \in \mathcal{S}_K^+(\rho)$ known as the *overlap matrix*, for reasons which will become clear later on. The associated posterior is

$$\mathbb{P}(u|\widetilde{Y}_0, V) = \frac{1}{\mathcal{Z}_{P_{\text{out}}}} \frac{e^{-\frac{1}{2} u^\intercal u}}{(2\pi)^{K/2}} P_{\text{out}}(\widetilde{Y}_0|q^{1/2} V + (\rho - q)^{1/2} u), \tag{4.4}$$

and the free entropy reads this time $\Psi_{\text{out}}(q; \rho) \equiv \mathbb{E} \ln \mathcal{Z}_{P_{\text{out}}}$.

In Appendix C.2 (more precisely Lemmas C.2 and C.3), we prove regularity and convexity properties of ψ_{P_0} and Ψ_{out} which will be useful for our analysis.

The free entropy

As we know from Section 1.1, the central object to study optimal performances is the posterior distribution of the weights:

$$\mathbb{P}(\mathbf{w}|\mathbf{X}, \mathbf{Y}) = \frac{1}{\mathcal{Z}_n} \prod_{i=1}^{n} P_0(w_i) \prod_{\mu=1}^{m} P_{\text{out}}\Big(Y_\mu \Big| \Big\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{\mu i} w_{il} \Big\}_{l=1}^{K} \Big), \tag{4.5}$$

The expected free entropy is by definition $f_n \equiv (1/n)\mathbb{E} \ln \mathcal{Z}_n$. Recall that we consider the *thermodynamic* limit $n, m \to \infty$ with $m/n \to \alpha > 0$. The main theoretical result of this chapter is a rigorous proof of the free entropy formula predicted by the replica method, using a probabilistic interpolation method developed in [Gue03, Tal03, BM19a].

We define the *replica symmetric (RS) potential* as

$$f_{\text{RS}}(q, r) \equiv \psi_{P_0}(r) + \alpha \Psi_{\text{out}}(q; \rho) - \frac{1}{2} \text{Tr}(rq), \tag{4.6}$$

where $\alpha \equiv m/n$, and $\Psi_{\text{out}}(q; \rho)$ and $\psi_{P_0}(r)$ are the free entropies of the two simpler K-dimensional estimation problems (4.3) and (4.4). We moreover assume the following:

H.1 The prior P_0 has bounded support in $\mathbb{R}^K$. Recall that $\rho = \mathbb{E}_{P_0}[W_0 W_0^\intercal]$.

H.2 The activation $\varphi_{\text{out}} : \mathbb{R}^K \times \mathbb{R} \to \mathbb{R}$ is a bounded $\mathcal{C}^2$ function with bounded first and second derivatives w.r.t. its first argument.

H.3 $X_{\mu i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

Theorem 4.1 (*Replica formula*)

Suppose H.1, H.2 and H.3. The asymptotic free entropy is:

$$\lim_{n\to\infty} f_n \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\ln \mathcal{Z}_n = \sup_{r\in\mathcal{S}_K^+} \inf_{q\in\mathcal{S}_K^+(\rho)} f_{\rm RS}(q,r)\,.$$

Theorem 4.1 is an extension of the results of [BKM+19] which studied usual generalized linear models (GLMs), i.e. the case $K = 1$ of this chapter. Our proof is based on an *adaptive interpolation method* developed in [BM19a], and is outlined in Section 4.4. We conclude the presentation of the theorem by two remarks.

Remark 4.1 (*Regularization*)

Theorem 4.1 actually stands under the addition of an (arbitrarily small) Gaussian regularization noise to the model (4.2), which thus becomes (with $\Delta > 0, Z_\mu \sim \mathcal{N}(0,1)$)

$$Y_\mu = \varphi_{\rm out}\Big(\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{\mu i}W^*_{il}\Big\}_{l=1}^{K}, A_\mu\Big) + \sqrt{\Delta}Z_\mu.$$

In general this regularizing noise is needed for the free entropy limit to exist in noiseless scenarios. Some exceptions exist, such as the case of discrete outputs (e.g. in the committee machine of eq. (4.1)), for which we can safely take $\Delta = 0$. The reader can refer to [BKM+19] for more details on this technical point.

Remark 4.2 (*Relaxing the hypotheses*)

Following similar approximation arguments as in [BKM+19], H.1 can be relaxed to the existence of the second moment of the prior, thus covering e.g. the Gaussian prior case. H.2 can moreover be dropped (so that our results include the committee machine (4.1)), and H.3 can be extended to $\mathbf{X}$ having i.i.d. entries of zero mean, unit variance and finite third moment.

Optimal generalization error

As we know, in the Bayes-optimal setting the estimator $\hat{\mathbf{W}}$ that minimizes the mean-square error with the ground-truth $\mathbf{W}^*$ is given by the expected mean of the posterior distribution of eq. (4.5). Denoting q^* the extremizer in the replica formula of Theorem 4.1, we expect from the replica method that with high probability over the quenched variables, $\hat{\mathbf{W}}^\intercal\mathbf{W}^*/n \xrightarrow{\rm p} q^*$. This phenomenon is known as *overlap concentration*, and will be discussed in the proof of Theorem 4.1.

From the *overlap* matrix q^*, one can compute the Bayes-optimal generalization error when the student tries to classify a new, yet unseen, sample $\mathbf{X}_{\rm new} \in \mathbb{R}^n$. Indeed, the estimator of the new label $\hat{Y}_{\rm new}$ that minimizes the mean-squared error with the true label is given by the posterior mean of $\varphi_{\rm out}(\mathbf{X}_{\rm new}^\intercal\mathbf{w})$. The *Bayes-optimal generalization error* is then

$$\frac{1}{2}\mathbb{E}_{\mathbf{X},\mathbf{W}^*}\left\{\Big(\mathbb{E}_{\mathbf{w}|\mathbf{X},Y}\left[\varphi_{\rm out}(\mathbf{X}_{\rm new}^\intercal\mathbf{w})\right] - \varphi_{\rm out}(\mathbf{X}_{\rm new}^\intercal\mathbf{W}^*)\Big)^2\right\} \xrightarrow[n\to\infty]{} \epsilon_g(q^*), \tag{4.7}$$

where $\mathbf{w}$ is distributed according to the posterior measure of eq. (4.5)[1]. In particular, when the distribution of $\mathbf{X}$ is rotationally invariant, the asymptotic generalization error only depends on the overlap matrix $\mathbf{w}^\intercal\mathbf{W}^*/n \xrightarrow{\rm p} q^*$. In specific cases (e.g. when $K = 2$ or in the limit $K \to \infty$ as we will see), one can derive explicit formulas for the function $\epsilon_g(q^\star)$.

[1] The Bayes-optimal error differs from the *Gibbs generalization error* by a factor 2, as detailed in Appendix D.1.

In Appendix D.1 we give more details on possible definitions of the generalization error, and derive its expression for $K = 2$. The analytical treatment of the limit $K \to \infty$ in $\epsilon_g(q^*)$ is detailed in Appendix D.2, which is devoted to this limit.

4.3 Investigating computational-to-statistical gaps

4.3.1 Approximate Message-Passing

In order to investigate optimal algorithmic performances in this model, we pick the appropriate tool from our toolbox. As we discussed in Section 1.4, the natural candidate to design an algorithm that is optimal among a very wide class of polynomial-time algorithms is *Approximate Message-Passing* (AMP). While AMP was initially developed to solve random instances of generalized linear models with Gaussian data matrices [DMM09, Ran11], here we adapt the algorithm for the committee machine, and more generically for models described by eq. (4.2). As in generalized linear models, in the Bayes-optimal setting AMP is strongly conjectured to perform the best among all polynomial algorithms (and shown to be optimal among the large class of general first order methods [CMW20]), in terms of achieved overlap. It thus gives us a tool to evaluate both the intrinsic algorithmic hardness of learning and the performance of existing algorithms with respect to the optimal one.

Algorithm 3: Approximate Message-Passing iterations for the committee machine

Input : A vector $\mathbf{Y} \in \mathbb{R}^m$, a data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$
Output: An estimator $\hat{\mathbf{W}} \in \mathbb{R}^{n \times K}$
Initialize variables randomly;
while $\| \hat{\boldsymbol{W}}^t - \hat{\boldsymbol{W}}^{t-1} \| > \epsilon$ **do**
  Update of the mean $\omega_\mu \in \mathbb{R}^K$ and covariance $V_\mu \in \mathcal{S}_K^+$:
  $\omega_\mu^t = \sum_{i=1}^n \left(\frac{X_{\mu i}}{\sqrt{n}} \hat{W}_i^t - \frac{X_{\mu i}^2}{n} (\Sigma_i^{t-1})^{-1} \hat{C}_i^t \Sigma_i^{t-1} g_{\text{out},\mu}^{t-1}\right) \quad | \quad V_\mu^t = \sum_{i=1}^n \frac{X_{\mu i}^2}{n} \hat{C}_i^t$;
  Update of $g_{\text{out},\mu} \in \mathbb{R}^K$ and $\partial_\omega g_{\text{out},\mu} \in \mathcal{S}_K^+$:
  $g_{\text{out},\mu}^t = g_{\text{out}}(\omega_\mu^t, Y_\mu, V_\mu^t) \quad | \quad \partial_\omega g_{\text{out},\mu}^t = \partial_\omega g_{\text{out}}(\omega_\mu^t, Y_\mu, V_\mu^t)$;
  Update of the mean $T_i \in \mathbb{R}^K$ and covariance $\Sigma_i \in \mathcal{S}_K^+$:
  $T_i^t = \Sigma_i^t \Big(\sum_{\mu=1}^m \frac{X_{\mu i}}{\sqrt{n}} g_{\text{out},\mu}^t - \frac{X_{\mu i}^2}{n} \partial_\omega g_{\text{out},\mu}^t \hat{W}_i^t \Big) \quad | \quad \Sigma_i^t = -\Big(\sum_{\mu=1}^m \frac{X_{\mu i}^2}{n} \partial_\omega g_{\text{out},\mu}^t \Big)^{-1}$;
  Update of the estimated marginals $\hat{W}_i \in \mathbb{R}^K$ and $\hat{C}_i \in \mathcal{S}_K^+$:
  $\hat{W}_i^{t+1} = f_w(\Sigma_i^t, T_i^t) \quad | \quad \hat{C}_i^{t+1} = f_c(\Sigma_i^t, T_i^t)$;
  $t = t + 1$;
end

The AMP algorithm for generalized committee machines is summarized in Algorithm 3. The update functions used in this algorithm are defined as:

- The functions $f_w(\Sigma, T)$ and $f_c(\Sigma, T)$ are respectively the mean and variance under the posterior distribution of the auxiliary model of eq. (4.3), when $r \to \Sigma^{-1}$ and $Y_0 \to \Sigma^{1/2} T$.
- $g_{\text{out}}(\omega_\mu, Y_\mu, V_\mu)$ is related to the mean of posterior of the second auxiliary model, i.e. eq. (4.4). More precisely one can define it as:

$$g_{\text{out}}(\omega, y, V) \equiv \frac{1}{V} \frac{\int_{\mathbb{R}^K} (z - \omega) e^{-\frac{1}{2}(z-\omega)^\intercal V^{-1}(z-\omega)} P_{\text{out}}(y|z) \mathrm{d}z}{\int_{\mathbb{R}^K} e^{-\frac{1}{2}(z-\omega)^\intercal V^{-1}(z-\omega)} P_{\text{out}}(y|z) \mathrm{d}z}. \tag{4.8}$$

The detailed derivation of Algorithm 3 from loopy belief propagation is quite tedious. While we do not detail it in this thesis, the interested reader may refer to [AMB+19]. After convergence, $\hat{\mathbf{W}}$ estimates the weights of the teacher neural network. A demonstration code of the algorithm is available on GitHub [AMB+18].

Let us briefly recall some properties of AMP that we highlighted in Section 1.4. A major strength of AMP is that its performance can be tracked rigorously in the asymptotic limit $n \to \infty$, via a procedure known as *State Evolution* (SE) [JM13]. In the present model, SE allows to track the value of the overlap between the teacher weights $\mathbf{W}^*$ and the AMP estimate $\hat{\mathbf{W}}^t$, defined as $q^t_{\text{AMP}} \equiv \lim_{n\to\infty}(\hat{\mathbf{W}}^t)^\intercal \mathbf{W}^*/n$, via the iteration of the following equations:

$$q^{t+1}_{\text{AMP}} = 2\nabla\psi_{P_0}(r^t_{\text{AMP}})\,, \qquad r^{t+1}_{\text{AMP}} = 2\alpha\nabla\Psi_{\text{out}}(q^t_{\text{AMP}};\rho)\,. \tag{4.9}$$

A crucial property of State Evolution (that is trivial to derive from definition) is that its fixed points exactly correspond to the critical points of the replica-symmetric potential of eq. (4.6). This allows to treat both the information-theoretic optimal performance (via Theorem 4.1) and the AMP performance (via eq. (4.9)) in a single framework. We will illustrate the power of this correspondence in the following Section 4.3.2.

On the convergence of the algorithm – In the large n limit, and if all functions are computed without errors, the algorithm is guaranteed to converge. This is a consequence of the Bayes-optimal setting, as detailed in [AMB+19]. In practice, of course, n is finite and K-dimensional integrals need to be approximated. In that case convergence is not guaranteed, but is robustly achieved in all the cases presented in this chapter. We also expect (by experience with the single layer case) that if the input-data matrix $\mathbf{X}$ has more structure than simply i.i.d. then the AMP described here would encounter convergence issues. Such issues could however be fixed by moving to refined variants of the algorithm, such as Vector Approximate Message-Passing (VAMP) [RSF17], cf. Section 1.4 for an introduction to the VAMP algorithm. Studying a "VAMP"-like algorithm in the context of the committee machine with structured data is however not in the scope of this chapter.

4.3.2 From two to more hidden neurons, and the specialization transition

A network with two hidden neurons

Let us now discuss how the above results (i.e. Theorem 4.1 and the AMP algorithm) can be used to study optimal learning in a simple (yet non-trivial) two-layers neural network with two hidden neurons. Concretely, we consider the following special case of Model 4.1[2]:

$$Y_\mu = \text{sign}\Big[\text{sign}\Big(\sum_{i=1}^n X_{\mu i} W^*_{i1}\Big) + \text{sign}\Big(\sum_{i=1}^n X_{\mu i} W^*_{i2}\Big)\Big].$$

In Fig. 4.2 we illustrate our results for this model. In the left panel the weights are Gaussian (i.e. $P_0 = \mathcal{N}(0,1)$), while in the right panel they are binary (or Ising/Rademacher, i.e. $P_0 = (\delta_{-1} + \delta_1)/2$). We plot in Fig. 4.2 several quantities:

- In black we show the generalization error. Full lines are obtained as global maximizers of the replica potential of eq. (4.6), while dashed lines are obtained from the fixed point of the State Evolution (SE) of the AMP algorithm (i.e. eq. (4.9)). Note that it corresponds to the *local maximizer* of the replica-symmetric potential of eq. (4.6), when iterating the SE equations starting at $q = 0$. Dots are finite-size simulations of the AMP algorithm.

[2] We adopt the convention $\text{sign}(0) = 0$.

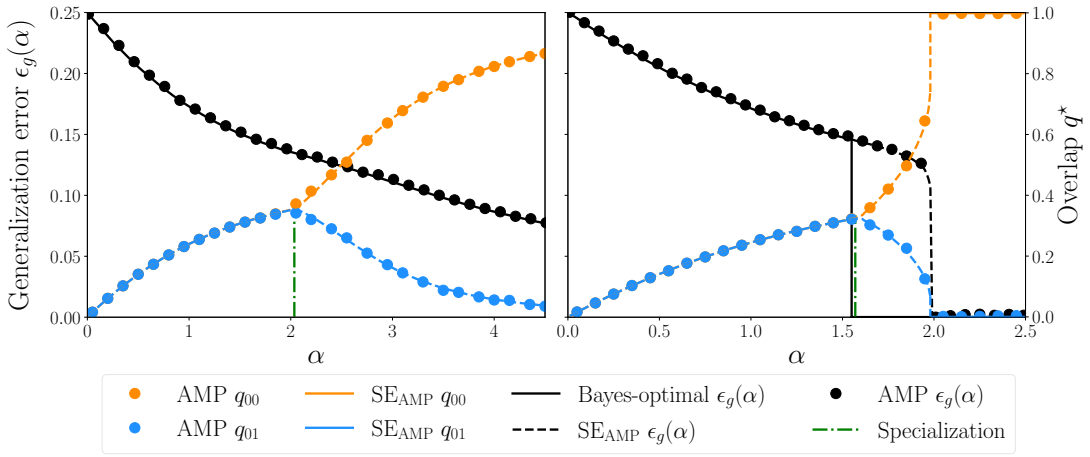


FIGURE 4.2: Generalization error and overlap matrix elements for a committee machine with two hidden neurons ($K = 2$) with Gaussian weights (left) and binary/Rademacher weights (right), as a function of the sampling ratio $\alpha = m/n$. q_{00} and q_{01} denote diagonal and off-diagonal overlaps, and their values are given by the labels on the far-right of the figure.

- In blue and orange we show the diagonal and off-diagonal element of the matrix overlap q achieved by AMP. Solid lines are obtained again using the SE equations (4.9), while dots are simulations of AMP. Finally, in green dashed-dot line we show the *specialization* point at which the diagonal and off-diagonal overlaps start differing.

We ran the AMP algorithm with $n = 10^4$ until convergence, and averaged over 10 instances with different random initial conditions. As expected by our theory we observe excellent agreement between the SE and AMP. Let us now comment further on our results.

Specialization – In both panels of Fig. 4.2 we observe the so-called *specialization* phase transition. Indeed eq. (4.9) has two types of fixed points: a *non-specialized* fixed point where every matrix element of the $K \times K$ order parameter q is the same (so that both hidden neurons learn the same function), and a *specialized* fixed point where the diagonal elements of the order parameter are different from the non-diagonal ones. We checked for other types of fixed points for $K = 2$ (one where the two diagonal elements are not the same), but have not found any. In terms of weight-learning, this means for the non-specialized fixed point that the estimators for both $\mathbf{W}_1$ and $\mathbf{W}_2$ are the same, whereas in the specialized fixed point the estimators of the weights corresponding to the two hidden neurons are different: the network "figured out" that the data is better described by a model that is not linearly separable. The specialized fixed point is associated with lower error than the non-specialized one (as one can see in Fig. 4.2). The existence of such a phase transition was previously discussed in the statistical physics literature, see e.g. [SH92, SS95]. Interestingly, one can show that specialization is absent (for arbitrary K) when the activations are linear, as detailed in [AMB+19]. The non-linearity of the activation functions is therefore an essential ingredient in order to observe a specialization transition.

Gaussian weights – For Gaussian weights (Fig. 4.2 left), the specialization phase transition arises continuously at $\alpha^G_{\text{spec}}(K = 2) \simeq 2.04$. This means that for $\alpha < \alpha^G_{\text{spec}}(K = 2)$ the number of samples is too small, and the student network is not able to learn that two different teacher-vectors $\mathbf{W}_1$ and $\mathbf{W}_2$ were used to generate the observed labels. For $\alpha > \alpha^G_{\text{spec}}(K = 2)$, however, it is able to distinguish the two different weight-vectors and the generalization error decreases fast to low values.

Binary weights – The right part of Fig. 4.2 depicts the case of binary weights. We observe two phase transitions in the performance of AMP in this case: (a) the specialization phase

transition at $\alpha^B_{\rm spec}(K=2)\simeq 1.58$, and for slightly larger sample complexity a transition towards *perfect generalization* (beyond which the generalization error is asymptotically zero) at $\alpha^B_{\rm perf}(K=2)\simeq 1.99$. The binary case with $K=2$ differs from the Gaussian one in the fact that perfect generalization is achievable at finite α. While the specialization transition is continuous here, the error has a discontinuity at the transition of perfect generalization. This discontinuity is associated with a 1st order phase transition (in the physics nomenclature), leading to a gap between algorithmic performance and information-theoretically optimal performance. To quantify said optimal performance we need to evaluate the global maximum of the replica free entropy (not the local one reached by the state evolution). In doing so that we discover that information-theoretically there is a single discontinuous phase transition towards perfect generalization at $\alpha^B_{\rm IT}(K=2)\simeq 1.54$.

Computational-to-statistical gap – While the information-theoretic and specialization phase transitions were partially discussed in the physics literature on the committee machine [SH92, SH93, SST92, WRB93], the gap between the information-theoretic performance and the performance of AMP —that is conjectured to be optimal among polynomial-time algorithms— was not previously discussed for such neural networks. Indeed, even its understanding in simpler models than those discussed here, such as the single layer case, is quite recent [DMM09, DJM13, ZK16]. The existence of similar gaps (or their absence) will also be among the core results of Chapters 5 and 6.

$K\gg 1$: when more is different

It becomes increasingly difficult to study the replica formula for larger values of K as it involves (at least) K-dimensional integrals. Quite interestingly, it is however possible to work out the solution of the replica formula in the large K limit (taken *after* the large n limit). Indeed, when the activations functions of all hidden units are identical, it is natural to look for solutions of the replica formula of the form $q=q_d\mathrm{I}_K+(q_a/K)\mathbf{1}_K\mathbf{1}_K^\intercal$, with the unit vector $\mathbf{1}_K=(1)_{l=1}^K$. Such solutions are what we call *committee symmetric*. To simplify the argument, we assume that the prior P_0 has covariance $\rho=\mathrm{I}_K$. Since both q and $\rho-q$ are positive, the committee-symmetry assumption implies that $0\leq q_d\leq 1$ and $0\leq q_a+q_d\leq 1$ in the large-K limit.

In Appendix D.2, we detail the corresponding large K expansion of the free entropy for the teacher-student scenario with Gaussian weights, and sign activation functions (the model of eq. (4.1)). While the information-theoretically optimal generalization error was previously computed [Sch93], we concentrated on the analysis of the performance of AMP by tracking the state evolution equations. In doing so, we unveil a large computational gap.

Description of our results – Our results at large K are presented in Fig. 4.3. In the right panel we show the fixed point values of the two overlaps $q_{00}=q_d+q_a/K$ and $q_{01}=q_a/K$. The resulting generalization error is plotted in the left panel. As shown in Appendix D.2 it can be written as $\epsilon_g=\arccos[2(q_a+\arcsin q_d)/\pi]/\pi$. The information-theoretic specialization transition arises for $\alpha=\Theta(K)$, so we define $\widetilde{\alpha}\equiv\alpha/K$. Contrary to the case $K=2$, specialization is here a 1st order phase transition, meaning that the specialization fixed point first appears at $\widetilde{\alpha}^G_{\rm spinodal}\simeq 7.17$ but the free entropy global extremizer remains the one of the non-specialized fixed point until $\widetilde{\alpha}^G_{\rm spec}\simeq 7.65$. This has interesting implications for the optimal generalization error that gets towards a plateau of value $\varepsilon_{\rm plateau}\simeq 0.28$ for $\widetilde{\alpha}<\widetilde{\alpha}^G_{\rm spec}$ and then jumps discontinuously down. For large $\widetilde{\alpha}$, the Bayes-optimal error decays asymptotically as $1/\widetilde{\alpha}$, i.e. $\epsilon_g^{\rm IT}(\widetilde{\alpha})=\Theta(1/\widetilde{\alpha})$.

A large computational gap – As we have already discussed throughout this thesis, AMP is conjectured to be optimal among a wide class of polynomial-time algorithms [CMW20], and thus analyzing its state evolution sheds light on possible computational gaps, that come hand in hand with 1st order phase transitions. In the regime $\alpha=\mathcal{O}(K^2)$ for large K, we find in Appendix D.2

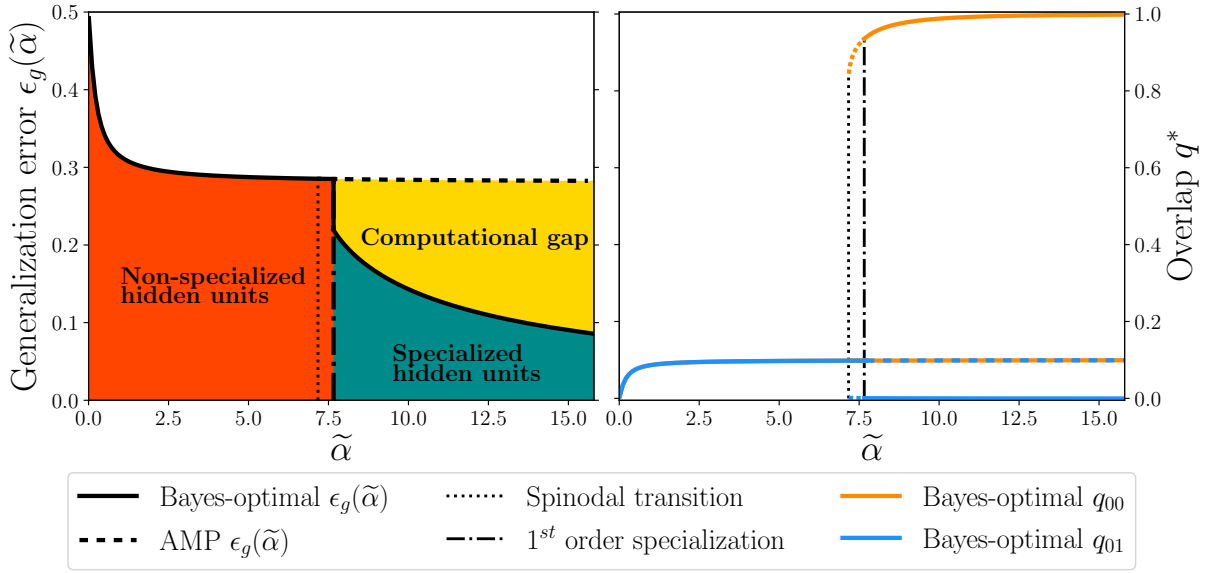


FIGURE 4.3: (Left) Bayes-optimal and AMP generalization errors and (right) diagonal and off-diagonal overlaps q_{00}, q_{01} for a committee machine with a large number of hidden neurons K and Gaussian weights, as a function of the rescaled parameter $\tilde{\alpha} = \alpha/K$. Solutions corresponding to global and local maxima of the replica free entropy are respectively represented with full and dashed lines. The dotted line marks the *spinodal transition* at $\widetilde{\alpha}^G_{\text{spinodal}} \simeq 7.17$, i.e. the apparition of a local minimum in the replica free entropy, associated to a solution with specialized hidden units. The dotted-dashed line shows the first order specialization transition at $\widetilde{\alpha}^G_{\text{spec}} \simeq 7.65$, at which the specialized fixed point becomes the global minimum. For $\widetilde{\alpha} < \widetilde{\alpha}^G_{\text{spec}}$, AMP reaches the Bayes-optimal generalization error and overlaps, corresponding to a non-specialized solution. However, for $\widetilde{\alpha} > \widetilde{\alpha}^G_{\text{spec}}$, the AMP algorithm does not follow the optimal specialized solution and is stuck in the non-specialized solution plateau, represented with dashed lines. Hence it unveils a large computational gap (yellow area).

that the non-specialized fixed point is always stable, implying that AMP will not be able to give a lower generalization error than $\varepsilon_{\text{plateau}}$ corresponding to a non-specialized solution. However, information-theoretically a specialization transition arises in the scale $\alpha = \Theta(K)$: this implies a large gap between the performance that can be reached information-theoretically and the one reachable tractably, see the yellow area in Fig. 4.3. Such large computational gaps have been previously identified in a range of inference problems —most famously in the planted clique problem [DM15]— but the committee machine is the first model of a multi-layer neural network with realistic non-linearity that presents such a large gap.

4.4 Proof of the replica formula: adaptive interpolation

In this section, we sketch the proof of Theorem 4.1 by doing a small excursion closer to the realm of probability. We use an *adaptive interpolation method*, based on an original idea of Guerra [Gue03], and refined by Barbier and collaborators, see [BM19a] for a review on this technique. In this regard, the interested reader can also consider this section as an introduction to this powerful technique, in a quite involved context[3]. Details of proof will often be postponed to Appendix C, as we focus on describing the method.

[3]For applications of this method to simpler models, one can refer to [BM19b] for e.g. the spiked Wigner model, or to [BKM+19] for GLMs.

All along this section we assume H.1, H.2 and H.3, and all our statements will implicitly assume these hypotheses. We also define a few quantities for the remaining of this section:

- For $\mu = 1, \dots m$, let V_μ, U_μ^* be two vectors drawn from $\mathcal{N}(0, \mathrm{I}_K)$.
- Let $s_n \in (0, 1/2]$ an arbitrary sequence s.t. $s_n \underset{n\to\infty}{\to} 0$.
- Let $\mathcal{M}$ be the compact subset of positive definite matrices in $\mathcal{S}_K$ with all eigenvalues in the interval $[1, 2]$. In particular, for all $M \in (s_n\mathcal{M})$, we have $(2s_n\mathrm{I}_K - M) \in \mathcal{S}_K^+$.

4.4.1 Interpolating estimation problem

Let $\epsilon = (\epsilon_1, \epsilon_2) \in (s_n\mathcal{M})^2$. Let $q : [0,1] \to \mathcal{S}_K^+(\rho)$ and $r : [0,1] \to \mathcal{S}_K^+$ be two "interpolation functions" (that will eventually depend on ϵ), and

$$R_1(t) \equiv \epsilon_1 + \int_0^t r(v)dv\,, \qquad R_2(t) \equiv \epsilon_2 + \int_0^t q(v)dv\,. \tag{4.10}$$

For $t \in [0,1]$, define the K-dimensional vector (using the matrix square root):

$$S_{t,\mu} \equiv \sqrt{\frac{1-t}{n}} \sum_{i=1}^n X_{\mu i} W_i^* + \sqrt{R_2(t)}\, V_\mu + \sqrt{t\rho - R_2(t) + 2s_n\mathrm{I}_K}\, U_\mu^*. \tag{4.11}$$

We will interpolate between our original problem, and auxiliary problems related to the ones of eqs. (4.3),(4.4). More precisely, the interpolating estimation problem is given by the following observation model, with two types of t-dependent observations:

$$\begin{cases} Y_{t,\mu} \sim P_{\mathrm{out}}(\,\cdot\,|\,S_{t,\mu}), & 1 \le \mu \le m, \\ Y'_{t,i} = \sqrt{R_1(t)}\, W_i^* + Z'_i \in \mathbb{R}^K, & 1 \le i \le n, \end{cases} \tag{4.12}$$

where for each i, $Z'_i \sim \mathcal{N}(0, \mathrm{I}_K)$. Recall that in our notation the "$*$-variables" have to be retrieved, while other random variables are fixed (they are *quenched variables* in the statistical physics language). Define now $s_{t,\mu}$ by a similar expression to eq. (4.11):

$$s_{t,\mu} \equiv \sqrt{\frac{1-t}{n}} \sum_{i=1}^n X_{\mu i} w_i + \sqrt{R_2(t)}\, V_\mu + \sqrt{t\rho - R_2(t) + 2s_n\mathrm{I}_K}\, u_\mu. \tag{4.13}$$

The posterior of the interpolating problem is given by:

$$\mathbb{P}_{t,\epsilon}(\mathbf{w}, \mathbf{u}|\mathbf{Y}_t, \mathbf{Y}'_t, \mathbf{X}, \mathbf{V}) = \frac{1}{\mathcal{Z}_{n,\epsilon}(t)} \prod_{i=1}^n P_0(w_i) e^{-\frac{\|Y'_{t,i} - \sqrt{R_1(t)} w_i\|^2}{2}} \prod_{\mu=1}^m \frac{e^{-\frac{\|u_\mu\|^2}{2}}}{(2\pi)^{K/2}} P_{\mathrm{out}}(Y_{t,\mu}|s_{t,\mu}). \tag{4.14}$$

The free entropy "at time t" is by definition (recall that $\mathcal{D}$ is the standard Gaussian measure)

$$f_{n,\epsilon}(t) \equiv \frac{1}{n}\mathbb{E}\ln \int \mathcal{D}\mathbf{u} \prod_{i=1}^n \mathrm{d}w_i P_0(w_i) e^{-\frac{1}{2}\|Y'_{t,i} - \sqrt{R_1(t)} w_i\|^2} \prod_{\mu=1}^m P_{\mathrm{out}}(Y_{t,\mu}|s_{t,\mu}), \tag{4.15}$$

At $t = 0$, the free entropy of eq. (4.15) is very close to the free entropy f_n of the original problem (cf. the statement of Theorem 4.1), a difference arising from the presence of the small "perturbation" ϵ. We first show that this perturbation does not change the asymptotic free entropy:

Lemma 4.2 (*Perturbation of the free entropy*)

There exists a constant $C > 0$ such that for all $\epsilon \in (s_n\mathcal{M})^2$ we have

$$|f_{n,\epsilon}(0) - f_{n,\epsilon=(0,0)}(0)| \leq C s_n.$$

Proof of Lemma 4.2 – One can compute easily[4]:

$$\nabla_{\epsilon_1} f_{n,\epsilon}(0) = -\frac{1}{2}[\rho - \mathbb{E}\langle Q\rangle_{n,0,\epsilon}],$$

where the *overlap matrix* $Q \in \mathcal{S}_K$ is defined below by eq. (4.18). Therefore, $\|\nabla_{\epsilon_1} f_{n,\epsilon}(0)\|_F$ is bounded (using H.1). We define $u_y(x) \equiv \ln P_{\text{out}}(y|x)$. We can also compute (by calculations very similar to the ones used in the proof of the following Proposition 4.3, so that we leave them for the reader):

$$\nabla_{\epsilon_2} f_{n,\epsilon}(0) = \frac{1}{2n}\sum_{\mu=1}^{m} \mathbb{E}\Big[\nabla u_{Y_{t,\mu}}(S_{t,\mu})\Big\langle \nabla u_{Y_{t,\mu}}(s_{t,\mu})\Big\rangle_{n,0,\epsilon}\Big].$$

Note that the r.h.s. of the above equation is symmetric by the Nishimori Proposition 1.1, and it is bounded by Hypothesis H.2. By the mean value theorem we obtain then directly that (for some constants C_1 and C_2) $|f_{n,\epsilon}(0) - f_{n,\epsilon=(0,0)}(0)| \leq \|\nabla_{\epsilon_1} f_{n,\epsilon}(0)\|_F \|\epsilon_1\|_F + \|\nabla_{\epsilon_2} f_{n,\epsilon}(0)\|_F \|\epsilon_2\|_F \leq C_1 \max_i \|\epsilon_i\| \leq C_2 s_n$. □

Let us now precisely relate the extremal values $\{f_{n,\epsilon}(0),\, f_{n,\epsilon}(1)\}$ to our original problem. Using Lemma 4.2 , and continuity and boundedness properties of ψ_{P_0} and Ψ_{out} stated in Lemmas C.2 and C.3, one can easily verify from the very definition of $f_{n,\epsilon}(t)$:

$$\begin{cases} f_{n,\epsilon}(0) & = f_n - \dfrac{K}{2} + \mathcal{O}_n(1) \\ f_{n,\epsilon}(1) & = \psi_{P_0}\Big(\displaystyle\int_0^1 r(t)\mathrm{d}t\Big) + \alpha\Psi_{\text{out}}\Big(\displaystyle\int_0^1 q(t)\mathrm{d}t;\rho\Big) - \dfrac{1}{2}\displaystyle\int_0^1 \mathrm{Tr}[\rho\, r(t)]\mathrm{d}t - \dfrac{K}{2} + \mathcal{O}_n(1)\,. \end{cases} \tag{4.16}$$

Here $\mathcal{O}_n(1)$ is meant uniformly in t, q, r, ϵ.

4.4.2 Overlap concentration and fundamental sum rule

Notice from (4.16) that at $t = 1$ the interpolating estimation problem constructs part of the RS potential (4.6), while at $t = 0$ it is the free entropy of the original model (4.2) (up to a constant). In order to compare these boundary values, we use the identity

$$f_n = f_{n,\epsilon}(0) + \frac{K}{2} + \mathcal{O}_n(1) = f_{n,\epsilon}(1) - \int_0^1 \frac{\mathrm{d}f_{n,\epsilon}(t)}{\mathrm{d}t}\mathrm{d}t + \frac{K}{2} + \mathcal{O}_n(1). \tag{4.17}$$

The next obvious step is therefore to compute the free entropy variation along the interpolation path. This is summarized in the following Lemma, which is proven in Appendix C.4:

[4] or directly obtain by the I-MMSE formula for vector channels [RPD18]

Proposition 4.3 (*Free entropy variation*)

Denote by $\langle - \rangle_{n,t,\epsilon}$ the (Gibbs) expectation w.r.t. the posterior $\mathbb{P}_{t,\epsilon}$ given by (4.14). Set $u_y(x) \equiv \ln P_{\text{out}}(y|x)$. Then for all $t \in [0,1]$ we have

$$\frac{df_{n,\epsilon}(t)}{dt} = -\frac{1}{2}\mathbb{E}\Big\langle \mathrm{Tr}\Big[\Big(\frac{1}{n}\sum_{\mu=1}^{m} \nabla u_{Y_{t,\mu}}(s_{t,\mu}) \nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal - r(t)\Big)(Q - q(t))\Big]\Big\rangle_{n,t,\epsilon} + \frac{1}{2}\mathrm{Tr}[r(t)(q(t)-\rho)] + \mathcal{O}_n(1).$$

Again $\mathcal{O}_n(1) \to 0$ is meant uniformly in t, q, r, ϵ. The $K \times K$ *overlap matrix* Q is defined as

$$Q_{ll'} \equiv \frac{1}{n}\sum_{i=1}^{n} W_{il}^* w_{il'} \,. \tag{4.18}$$

We will then plug the expression of Proposition 4.3 into eq. (4.17). In order to simplify it we need the following crucial proposition, which states that the overlap concentrates. As we have discussed in Chapter 1, this property is equivalent to what we know as *replica symmetry*.

Proposition 4.4 (*Overlap concentration - Replica symmetry*)

Assume that for any $t \in (0,1)$ the transformation $\epsilon \in (s_n\mathcal{M})^2 \mapsto (R_1(t,\epsilon), R_2(t,\epsilon))$ is a $\mathcal{C}^1$ diffeomorphism with a Jacobian determinant greater or equal to 1. Then one can find a sequence s_n going to 0 slowly enough such that there exists a constant $C > 0$ depending only on φ_{out}, P_0, K and α, and a constant $\gamma > 0$ such that:

$$\frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2}\int_{(s_n\mathcal{M})^2} \mathrm{d}\epsilon \int_0^1 \mathrm{d}t\, \mathbb{E}\big\langle \|Q - \mathbb{E}\langle Q\rangle_{n,t,\epsilon}\|_{\mathrm{F}}^2 \big\rangle_{n,t,\epsilon} \le \frac{C}{n^\gamma}.$$

Proof of Proposition 4.4 – The proof of this concentration result can be directly taken from [Bar19], which is entirely devoted to proving overlap matrix concentration. Using the results of [Bar19] is straightforward assuming that $\epsilon \mapsto R(t,\epsilon)$ is a $\mathcal{C}^1$ diffeomorphism with a Jacobian determinant greater or equal to 1. The reader can also refer to [AMB+19] for more details. □

Combining eq. (4.17) with Propositions 4.3 and 4.4, we can deduce the following *fundamental sum rule* which is at the core of the proof:

Proposition 4.5 (*Fundamental sum rule*)

Assume that the interpolation functions r and q are such that the map $\epsilon = (\epsilon_1, \epsilon_2) \mapsto R(t,\epsilon) = (R_1(t,\epsilon), R_2(t,\epsilon))$ given by (4.10) is a $\mathcal{C}^1$ diffeomorphism whose Jacobian determinant is greater or equal to 1. Assume that for all $t \in [0,1]$ and $\epsilon \in (s_n\mathcal{M})^2$ we have $q(t) = q(t,\epsilon) = \mathbb{E}\langle Q\rangle_{n,t,\epsilon} \in \mathcal{S}_K^+(\rho)$. Then

$$f_n = \frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2}\int_{(s_n\mathcal{M})^2} \mathrm{d}\epsilon \Big\{ \psi_{P_0}\Big(\int_0^1 r(t)\mathrm{d}t\Big) + \alpha\Psi_{\text{out}}\Big(\int_0^1 q(t,\epsilon)\mathrm{d}t;\rho\Big) - \frac{1}{2}\int_0^1 \mathrm{Tr}[q(t,\epsilon)r(t)]\mathrm{d}t\Big\} + \mathcal{O}_n(1)\,.$$

In Proposition 4.5 we already see what we mean by *adaptive* interpolation: indeed, in order to apply this identity, we need $q(t)$, which is a parameter of the model, to be equal to the (averaged) overlap of said model: this will create a self-consistent condition on $q(t)$.

Proof of Proposition 4.5 – Let us denote $V_n \equiv \mathrm{Vol}(s_n\mathcal{M})^2$. In the following, the integral over ϵ is always over $(s_n\mathcal{M})^2$. Consider the first term, i.e. the Gibbs bracket, in the free entropy

derivative given by Proposition 4.3. By the Cauchy-Schwarz inequality we have

$$\Big(\mathbb{E}\Big\langle \mathrm{Tr}\Big[\Big(\frac{1}{n}\sum_{\mu=1}^{m}\nabla u_{Y_{t,\mu}}(s_{t,\mu})\nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal - r(t)\Big)\big(Q-q(t)\big)\Big]\Big\rangle_{n,t,\epsilon}\Big)^2$$
$$\leq \frac{1}{V_n}\int \mathrm{d}\epsilon \int_0^1 \mathrm{d}t\, \mathbb{E}\Big\langle \Big\| \frac{1}{n}\sum_{\mu=1}^{m}\nabla u_{Y_{t,\mu}}(s_{t,\mu})\nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal - r(t)\Big\|_F^2\Big\rangle_{n,t,\epsilon} \times \frac{1}{V_n}\int \mathrm{d}\epsilon \int_0^1 \mathrm{d}t\, \mathbb{E}\langle \|Q-q(t)\|_F^2\rangle_{n,t,\epsilon}.$$

The second term of this product is bounded by $Cn^{-\gamma}$ by Proposition 4.4, since we assumed that for all $\epsilon \in (s_n\mathcal{M})^2$ and all $t\in[0,1]$ we have $q(t)=q(t,\epsilon)=\mathbb{E}\langle Q\rangle_{n,t,\epsilon}$. The first term can itself be shown to be bounded by some constant C' that only depend on φ_{out} and α. This is shown in Lemma C.5 in Appendix C.5. In the end, applying Proposition 4.3 we obtain

$$\frac{1}{V_n}\int \mathrm{d}\epsilon \int_0^1 \frac{\mathrm{d}f_{n,\epsilon}(t)}{\mathrm{d}t}\mathrm{d}t = \frac{1}{2V_n}\int \mathrm{d}\epsilon \int_0^1 \mathrm{Tr}\big[q(t,\epsilon)r(t) - r(t)\rho\big]\mathrm{d}t + \mathcal{o}_n(1). \tag{4.19}$$

Again, the term $\mathcal{o}_n(1)$ is uniform w.r.t. to the choice of q and r. When replacing eq. (4.19) in eq. (4.17) and combining it with eq. (4.16) we reach the claimed identity. □

4.4.3 A technical lemma and an assumption

We give here a technical lemma used in the rest of the proof, and which allows us to detail the assumption on which we rely to prove Theorem 4.1.

Lemma 4.6 (*Technical lemma*)

Recall that $\mathbb{E}\langle Q\rangle_{n,t,\epsilon}$ is a function of $(n,t,R(t,\epsilon))$. Let $F_n^{(1)}(t,R(t,\epsilon)) \equiv 2\alpha\nabla\Psi_{\text{out}}(\mathbb{E}\langle Q\rangle_{n,t,\epsilon})$ and $F_n^{(2)}(t,R(t,\epsilon)) \equiv \mathbb{E}\langle Q\rangle_{n,t,\epsilon}$. $F_n \equiv (F_n^{(1)}, F_n^{(2)})$ is defined on the set:

$$D_n = \Big\{(t,r_1,r_2)\in[0,1]\times\mathcal{S}_K^+\times\mathcal{S}_K^+ \Big| (\rho t - r_2 + 2s_n I_K)\in\mathcal{S}_K^+\Big\}.$$

F_n is a continuous function from D_n to $\mathcal{S}_K^+\times\mathcal{S}_K^+(\rho)$. Moreover, F_n admits partial derivatives with respect to R_1 and R_2 on the interior of D_n. For every $(t,R(t,\epsilon))$ for which they are defined, they satisfy:

$$\sum_{l\leq l'}^{K}\frac{\partial (F_n^{(1)})_{ll'}}{\partial (R_1)_{ll'}} \geq 0. \tag{4.20}$$

Proof of Lemma 4.6 – The fact that the image domain of F_n is $\mathcal{S}_K^+\times\mathcal{S}_K^+(\rho)$ is a consequence of Lemma C.1. The continuity and differentiability of F_n follow from standard theorems of continuity and derivation under the integral sign (recall that we are working at finite n). Indeed, the domination hypotheses are easily satisfied since we work under H.1 and H.2. Let us now prove (4.20). We write the formal differential of $F_n^{(1)}$ with respect to R_1 as $\mathcal{D}_{R_1}F_n^{(1)}$, which is a 4-tensor, and our goal is to prove that $\mathrm{Tr}[\mathcal{D}_{R_1}F_n^{(1)}]\geq 0$, the trace of a 4-tensor $A_{(ij)(kl)}$ over S_K being $\mathrm{Tr}[A]\equiv\sum_{i\leq j}A_{(ij)(ij)}$. Then one can write $\mathrm{Tr}[\mathcal{D}_{R_1}F_n^{(1)}] = 2\alpha\mathrm{Tr}[\nabla\nabla^\intercal\Psi_{\text{out}}(\mathbb{E}\langle Q\rangle_{n,t,\epsilon})\times \nabla_{R_1}\mathbb{E}\langle Q\rangle_{n,t,\epsilon}]$. We know from Lemmas C.1 and C.3 that $\nabla\nabla^\intercal\Psi_{\text{out}}(\mathbb{E}\langle Q\rangle_{n,t,\epsilon})$ is a positive linear operator over $\mathcal{S}_K$. Moreover, it is a known result that the derivative $\nabla_{R_1}\mathbb{E}\langle Q\rangle_{n,t,\epsilon}$ is also positive, since R_1 is the "matrix snr" of a *linear* channel (see e.g. [RPD18]). Since the product of two symmetric positive matrices (here linear operators on $\mathcal{S}_K$) has always positive trace, this shows that $\mathrm{Tr}[\mathcal{D}_{R_1}F_n^{(1)}]\geq 0$. □

We can now state a technical assumption on which we rely, and which essentially allows us to derive that the map $\epsilon \mapsto R(t,\epsilon)$ is a $\mathcal{C}^1$ diffeomorphism with a Jacobian determinant greater or equal to 1 as will become clear in the next section:

Hypothesis 4.1 (*Technical assumption*)

With the notations of Lemma 4.6,

$$\sum_{l\leq l'}^{K} \frac{\partial (F_n^{(2)})_{ll'}}{\partial (R_2)_{ll'}} \geq 0.$$

An unnecessary hypothesis – Note that when [AMB+19] first came out, the proof was relying on the unproven Hypothesis 4.1. Since then, the work of Barbier & Reeves [BR20] has shown that this hypothesis is not necessary if one uses a slightly more involved interpolation path than the one of eq. (4.12), so that Theorem 4.1 stands without the need for Hypothesis 4.1. However, since this technicality significantly lengthens our proof without adding any new physical insight, in this thesis we chose to detail our original approach, assuming Hypothesis 4.1.

4.4.4 Matching bounds: adapting the interpolation path

We can now turn to proving the final result. We will use two different adaptive interpolation paths, to obtain respectively lower and upper bounds on the free entropy.

Proposition 4.7 (*Lower bound*)

Under Assumption 4.1, the free entropy associated to the posterior of eq. (4.5) verifies

$$\liminf_{n\to\infty} f_n \geq \sup_{r\in\mathcal{S}_K^+} \inf_{q\in\mathcal{S}_K^+(\rho)} f_{\rm RS}(q,r).$$

Proof of Proposition 4.7 – Choose first $r(t) = r \in \mathcal{S}_K^+$ a fixed matrix. Then $R(t) = (R_1(t), R_2(t))$ can be fixed as the solution to the first order Cauchy problem:

$$\frac{\mathrm{d}}{\mathrm{d}t} R_1(t) = r\,, \qquad \frac{\mathrm{d}}{\mathrm{d}t} R_2(t) = \mathbb{E}\langle Q\rangle_{n,t,\epsilon}\,, \qquad \text{and} \qquad R(0) = \epsilon\,. \tag{4.21}$$

We denote this (unique) solution $R(t,\epsilon) = (rt+\epsilon_1, \int_0^t q(v,\epsilon;r)\mathrm{d}v + \epsilon_2)$. It is possible to check that this ODE satisfies the hypotheses of the parametric Cauchy-Lipschitz theorem, and that by the Liouville formula the determinant $J_{n,\epsilon}(t)$ of the Jacobian of $\epsilon \mapsto R(t,\epsilon)$ satisfies (see Lemma C.4 for details on the Liouville formula)

$$J_{n,\epsilon}(t) = \exp\Big(\int_0^t \sum_{l\geq l'}^{K} \frac{\partial \mathbb{E}\langle Q_{ll'}\rangle_{n,s,\epsilon}}{\partial (R_2)_{ll'}}(s, R(s,\epsilon))\,\mathrm{d}s\Big) \overset{(a)}{\geq} 1, \tag{4.22}$$

in which (a) is a consequence of Assumption 4.1. Moreover by eq. (4.21), $q(t,\epsilon;r) = \mathbb{E}\langle Q\rangle_{n,t,\epsilon}$, which is in $\mathcal{S}_K^+$ by Lemma C.1. The fact that the map $\epsilon \mapsto R(t,\epsilon)$ is a $\mathcal{C}^1$ diffeomorphism is easily verified by its bijectivity (from the positivity of $J_{n,\epsilon}(t)$) combined with the local inversion theorem. All the assumptions of Proposition 4.5 are verified which then implies (recall the RS potential of eq. (4.6)):

$$f_n = \frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2} \int_{(s_n\mathcal{M})^2} \mathrm{d}\epsilon\, f_{\rm RS}\Big(\int_0^1 q(v,\epsilon;r)\mathrm{d}v, r\Big) + \mathcal{O}_n(1).$$

Since this is true for any $r \in \mathcal{S}_K^+$, this easily implies the lower bound. □

Proposition 4.8 (*Upper bound*)

Under Assumption 4.1, the free entropy associated to the posterior of eq. (4.5) verifies

$$\limsup_{n\to\infty} f_n \leq \sup_{r\in\mathcal{S}_K^+} \inf_{q\in\mathcal{S}_K^+(\rho)} f_{\rm RS}(q,r).$$

Proof of Proposition 4.8 – We now fix $R(t) = (R_1(t), R_2(t))$ as the solution $R(t,\epsilon) = (\int_0^t r(v,\epsilon)\mathrm{d}v + \epsilon_1, \int_0^t q(v,\epsilon)\mathrm{d}v + \epsilon_2)$ to the following Cauchy problem:

$$\frac{\mathrm{d}}{\mathrm{d}t}R_1(t) = 2\alpha\nabla\Psi_{P_{\rm out}}(\mathbb{E}\langle Q\rangle_{n,t,\epsilon})\,, \qquad \frac{\mathrm{d}}{\mathrm{d}t}R_2(t) = \mathbb{E}\langle Q\rangle_{n,t,\epsilon}\,, \qquad \text{and} \qquad R(0) = \epsilon.$$

We denote this equation as $\partial_t R(t) = F_n(t, R(t)), R(0) = \epsilon$. It is then possible to verify that $F_n(R(t),t)$ is a bounded $\mathcal{C}^1$ function of $R(t)$, and thus a direct application of the Cauchy-Lipschitz theorem implies that $R(t,\epsilon)$ is a $\mathcal{C}^1$ function of t and ϵ. The Liouville formula (Lemma C.4) for the Jacobian determinant of the map $\epsilon \in (s_n\mathcal{M})^2 \mapsto R(t,\epsilon)$ gives this time

$$J_{n,\epsilon}(t) = \exp\Big(\int_0^t \sum_{l\geq l'}^{K} \Big\{\frac{\partial(F_{n,1})_{ll'}}{\partial(R_1)_{ll'}}(s,R(s,\epsilon)) + \frac{\partial(F_{n,2})_{ll'}}{\partial(R_2)_{ll'}}(s,R(s,\epsilon))\Big\}\mathrm{d}s\Big) \overset{(a)}{\geq} 1\,. \tag{4.23}$$

Equality (a) follows again from the positivity of this sum of partials, see Lemma 4.6 and Assumption 4.1. Eq. (4.23) implies the bijectivity of $\epsilon \mapsto R(t,\epsilon)$ which, combined with the local inversion theorem, makes it a diffeomorphism. Since $\mathbb{E}\langle Q\rangle_{n,t,\epsilon}$ and $\rho - \mathbb{E}\langle Q\rangle_{n,t,\epsilon}$ are positive matrices (see Lemma C.1) we also have that $q(t,\epsilon) \in \mathcal{S}_K^+(\rho)$ and since by the differential equation we have $r(t,\epsilon) = 2\alpha\nabla\Psi_{\rm out}(q(t,\epsilon))$, we also have $r(t,\epsilon) \in \mathcal{S}_K^+$ by Lemma C.3. We have everything needed to apply Proposition 4.5 again which gives in this case

$$f_n = \frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2}\int \mathrm{d}\epsilon\Big\{\psi_{P_0}\Big(\int_0^1 r(v,\epsilon)\mathrm{d}v\Big) + \alpha\Psi_{\rm out}\Big(\int_0^1 q(v,\epsilon)\mathrm{d}v;\rho\Big) - \frac{1}{2}\mathrm{Tr}\int_0^1 q(v,\epsilon)r(v,\epsilon)\mathrm{d}v\Big\} + \mathcal{O}_n(1).$$

Then by convexity of ψ_{P_0} and $\Psi_{\rm out}$ (see Lemmas C.2 and C.3) and Jensen's inequality:

$$\begin{aligned} f_n &\leq \frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2}\int \mathrm{d}\epsilon\int_0^1 \mathrm{d}v\Big\{\psi_{P_0}(r(v,\epsilon)\mathrm{d}v) + \alpha\Psi_{\rm out}(q(v,\epsilon);\rho) - \frac{1}{2}\mathrm{Tr}[q(v,\epsilon)r(v,\epsilon)]\Big\} + \mathcal{O}_n(1), \\ &= \frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2}\int \mathrm{d}\epsilon\int_0^1 \mathrm{d}v\, f_{\rm RS}(q(v,\epsilon),r(v,\epsilon)) + \mathcal{O}_n(1)\,. \end{aligned}$$

We now remark that

$$f_{\rm RS}[q(v,\epsilon),r(v,\epsilon)] = \inf_{q\in\mathcal{S}_K^+(\rho)} f_{\rm RS}[q,r(v,\epsilon)].$$

Indeed, for every $r \in \mathcal{S}_K^+$, the function $g_r : q \in \mathcal{S}_K^+(\rho) \mapsto f_{\rm RS}(q,r)$ is convex (by Lemma C.3), and moreover $\nabla g_r(q) = \alpha\nabla\Psi_{\rm out}(q) - r/2$. Thus $\nabla g_{r(v,\epsilon)}(q(v,\epsilon)) = 0$ by definition of $r(v,\epsilon)$. Since $\mathcal{S}_K^+(\rho)$ is convex, the minimum of $g_{r(v,\epsilon)}(q)$ is necessarily achieved at $q = q(v,\epsilon)$. Therefore:

$$f_n \leq \frac{1}{\mathrm{Vol}(s_n\mathcal{M})^2}\int_{(s_n\mathcal{M})^2}\mathrm{d}\epsilon\int_0^1 \mathrm{d}v \inf_{q\in\mathcal{S}_K^+(\rho)} f_{\rm RS}(q,r(v,\epsilon)) + \mathcal{O}_n(1) \leq \sup_{r\in\mathcal{S}_K^+}\inf_{q\in\mathcal{S}_K^+(\rho)} f_{\rm RS}(q,r) + \mathcal{O}_n(1),$$

which concludes the proof of Proposition 4.8. □

Combining the two matching bounds of Propositions 4.7 and 4.8 ends the proof of Theorem 4.1.

Conclusion of Chapter 4

This chapter aimed at presenting how the different elements of the toolbox introduced in Chapter 1 can be combined to study optimal learning in a large class of two-layers neural networks.

One of its main contributions is the design of an AMP-type algorithm that is able to achieve Bayes-optimal learning in the high-dimensional limit. For the committee machine with a large number of hidden neurons we uncover regimes in which a low generalization error is information-theoretically achievable while the AMP algorithm fails to deliver it; strongly suggesting that no efficient algorithm exists for those cases, and unveiling a large computational gap. These *hard phases* are associated to first-order phase transitions in the solution of the model.

Importantly, we studied the *Bayes-optimal* setting in which the student network has the same architecture as the teacher. However, the replica method can be generalized to a case in which the teacher and students have different architectures, as was done for compressive sensing, see e.g. [KMS+12]. It is an interesting subject for future work to leverage the replica method to study how the hard phase evolves under e.g. over-parametrization. Note that since the writing of [AMB+19], works have studied the influence of under and over-parametrization in the committee machine on the performance of stochastic gradient descent, see e.g. [GAS+20].

Even though we focused in this chapter on a two-layers neural network, the analysis and algorithm can be extended to a multi-layer setting, following e.g. the lines of [MP92], as long as the number of layers as well as the number of hidden neurons in each layer is $\Theta_n(1)$, and as long as one only learns the first-layer weights. Indeed, such models still belong to the class of Model 4.1, so that Theorem 4.1 applies. The numerical evaluation of the phase diagram would be more challenging than the cases presented in this chapter, as multiple integrals would appear in the corresponding formulas. Going even further, one could study the case in which the weights of *all layers* are learned. A possible strategy to tackle this setting (still assuming a number $\Theta_n(1)$ of hidden neurons) would be to combine our AMP algorithm for the first layer with Expectation-Maximization procedures to learn the remaining layers. Similar ideas were already implemented in simpler settings than neural networks, see e.g. [KMS+12, KRFU14].

Let us make a final comment that will shed light on some of our motivation for the upcoming Part III. There exists a complex line of work that studies the loss-function landscape of neural networks. While a range of works show under various assumptions that spurious local minima are absent in neural networks, others show under different conditions that they do exist, see e.g. [SS18]. In the models we studied here, the regime of parameters that is hard for AMP must have spurious local minima, but the converse is not true in general[5]: this motivates a precise study of the loss landscape of neural networks, that will be the subject of Part III.

[5] Indeed it might be that there are spurious local minima, yet the AMP approach succeeds.

Chapter 5

Generative models, or how to exploit the structure in the data

"Data! Data! Data! ... I can't make bricks without clay!"

Sir Arthur Conan Doyle, Adventures of Sherlock Holmes (1892).

Disclaimer – While Chapter 4 illustrated well the power of our statistical physics toolbox to study optimal learning, it was restricted to i.i.d. data. In this chapter, we go further in this regard by investigating a crucial element to understand learning in practice: the structure in the data. In order to study theoretically such structure, we leverage *generative models*: more precisely, the data will be taken as the output of a (random) neural network. This will allow to model synthetic data sets that possess very non-trivial correlations arising from the structure of the generative model[1], and to understand how such structure can affect learning. We will focus on a rather simple inference problem to recover said data, namely *spiked matrix estimation*. This chapter is mainly based on the published work [ALM+20], and will illustrate the flexibility of the toolbox of Chapter 1, as we will adapt it to study the influence of data structure on the feasibility of learning. It also contains a detailed random matrix analysis in Section 5.4, which is inspired by the famous "BBP" transition [BBAP05].

5.1 Generative models for spiked matrix estimation

5.1.1 Introduction: exploiting data structure

Taking advantage of specific structures to enhance signal reconstruction is a central endeavour in modern signal processing. Notable technological advances - such as e.g. JPEG and MP3 compression - stem from the fact that images and sound admit a sparse representation in wavelet and Fourier bases. In a series of seminal works, see e.g. [CRT06], Candès, Romberg and Tao have shown that underparametrized linear systems can be inverted if the signal is assumed to be sparse. This opened the door for novel sub-Nyquist sampling strategies leveraged by sparsity which are at the heart of compressed sensing [Don06]. But interest in sparse representations reaches far beyond compressed sensing, and similar results have been derived for other signal processing tasks, such as sparse coding and sparse principal component analysis (PCA). Despite the remarkable success of these results, they broadly assume the latent sparse representation is given, thus relying on expert knowledge for signal pre-processing.

On the other hand, recent progress in deep learning has witnessed a surge of interest in neural network-based generative models. Opposed to sparsity, generative networks are trained to learn a latent representation of the structured signal. The expressiveness of neural networks allied with

[1] Understanding the influence of data structure in learning through synthetic generative models has recently received a burst of attention from some closes collaborators, see e.g. [GMKZ20, ALB+20, GLK+20, GRM+20].

the capacity to capture hierarchical representations led to impressive results in signal modelling, the most notable perhaps being Generative Adversarial Networks (GANs), which can be trained to generate realistic images of human faces [GPAM+14]. An important and natural question to ask is whether signals from generative models enjoy the same aforementioned interesting properties as sparse signals in reconstruction tasks. A series of recent results indeed suggest that the latent structure in generative models can be leveraged to improve signal reconstruction, see e.g. [TMC+16, BJPD17, HV18, HLV18, GMKZ20] among many other works. These suggest indeed that, in the words of [Vil18], "Generative models are the new sparsity".

This chapter is a further step in this direction: we analyze a class of random-neural generative priors in an unsupervised task: rank-one matrix factorization. Given a "data" matrix $\mathbf{Y} \in \mathbb{R}^{n\times p}$, the problem consists in finding two vectors (known as *spikes*) $\mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^p$ such that $\mathbf{Y}$ can be factorized as $\mathbf{Y} = \mathbf{u}\mathbf{v}^\intercal + \sqrt{\Delta}\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is an i.i.d. noise matrix of unit variance. This model is widely studied as a prototype for *principal component analysis* (PCA), since for small noise ($\Delta < 1$) and Gaussian spikes $\mathbf{u}, \mathbf{v}$, the optimal estimator is given by the leading principal component of $\mathbf{Y}$ [BBAP05]. Optimality relies on the assumption of unstructured spikes, and no longer holds if one of the spikes is sparse. In a similar spirit to compressed sensing, the investigation of sparse spikes in this model resulted into bespoke algorithms widely studied under the umbrella of sparse-PCA, see e.g. [JOB10].

An important conclusion of the aforementioned works is the existence of an algorithmic gap for sparse signal reconstruction: even if signal reconstruction is information-theoretically possible, no polynomial-time algorithm is known. For spiked-matrix factorization, this means that even though the best known sparse-PCA algorithm performs better than "vanilla" PCA, it doesn't reach the optimal threshold set by the theoretical Bayesian estimator. As we will show, this is in sharp contract to the class of neural generative models we study, for which we provide a polynomial time algorithm reaching the optimal theoretical performance. Our line of work has been continued by collaborators to study generative models in the phase retrieval problem, see [ALB+20].

5.1.2 Inference model: spiked matrix estimation

We will focus on the following two models, which are widely studied in the sparse-PCA literature [RF12, DM14a, LKZ15, PWBM16, BDM+16, Mio17, LM19] (note that they are as well very natural random matrix theory problems):

Model 5.1 (*Spiked Wigner model $vv^\intercal$*)

Consider an unknown vector (the *spike*) $\mathbf{v}^\star \in \mathbb{R}^p$ drawn from a distribution P_v on $\mathbb{R}^p$; we observe a matrix $\mathbf{Y} \in \mathbb{R}^{p\times p}$ with a symmetric noise term $\boldsymbol{\xi} \in \mathcal{S}_p$ and $\Delta > 0$:

$$\mathbf{Y} = \frac{1}{\sqrt{p}}\mathbf{v}^\star\mathbf{v}^{\star\intercal} + \sqrt{\Delta}\boldsymbol{\xi}, \tag{5.1}$$

where $\boldsymbol{\xi}/\sqrt{p} \sim \mathrm{GOE}(p)$, i.e. $\xi_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1+\delta_{ij})$ for $i \le j$. The aim is to find back the spike $\mathbf{v}^\star$ from $\mathbf{Y}$ (up to a global sign).

Model 5.2 (*Spiked Wishart/covariance model $\boldsymbol{uv}^\intercal$*)

Consider two unknown vectors $\mathbf{u}^\star \in \mathbb{R}^n$ and $\mathbf{v}^\star \in \mathbb{R}^p$ drawn from distributions P_u and P_v and let $\boldsymbol{\xi} \in \mathbb{R}^{n\times p}$ with $\xi_{\mu i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, and $\Delta > 0$. We observe

$$\mathbf{Y} = \frac{1}{\sqrt{p}}\mathbf{u}^\star\mathbf{v}^{\star\intercal} + \sqrt{\Delta}\boldsymbol{\xi}, \tag{5.2}$$

and the goal is to find back the spikes $\mathbf{u}^\star$ and $\mathbf{v}^\star$ from $\mathbf{Y} \in \mathbb{R}^{n\times p}$.

The high-dimensional limit that we consider (the *thermodynamic limit*) is $p, n \to \infty$ while $\tau \equiv n/p = \Theta(1)$, and the noise variance is $\Delta = \Theta(1)$. The prior P_v is representing the spike $\mathbf{v}^\star \in \mathbb{R}^p$ via a k-dimensional parametrization with $\alpha \equiv p/k = \Theta(1)$. Note that in sparse estimation, k is the number of non-zeros components of $\mathbf{v}^\star$, while in generative models k is the number of latent variables. As in Chapter 4, we assume a Bayes-optimal setting (introduced in Section 1.1): the priors and the noise $\Delta > 0$ are known to the observer.

5.1.3 Generative models for the data

Multivariate Gaussian prior

The simplest non-separable prior P_v that one can consider is the Gaussian model with a covariance matrix $\boldsymbol{\Sigma}$, that is $P_v = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. While this prior is not compressive, it captures some structure and can be simply estimated from data via the empirical covariance. This prior is very elementary, and we introduce it here as we will use it later to produce Fig. 5.5.

Multi-layer generative prior

To exploit the practically observed power of generative models, one would ideally consider models trained on datasets of examples of possible spikes, e.g. GANs, variational auto-encoders, or restricted Boltzmann machines. Such training, however, leads to strong correlations between the weights of the underlying neural networks, for which the theoretical part of the present chapter can not be applied readily. To allow for tractable theoretical analysis we focus on multi-layer generative models where all the weight matrices $\mathbf{W}^{(l)}$ $(1 \le l \le L)$ are fixed, independent, and generated i.i.d. from $\mathcal{N}(0,1)$. Let $\mathbf{v} \in \mathbb{R}^p$ be the output of such a generative model:

$$\mathbf{v} = \varphi^{(L)}\Big(\frac{1}{\sqrt{k_L}}\mathbf{W}^{(L)} \dots \varphi^{(1)}\Big(\frac{1}{\sqrt{k}}\mathbf{W}^{(1)}\mathbf{z}\Big)\Big), \tag{5.3}$$

with $\mathbf{z} \in \mathbb{R}^k$ a latent variable drawn from separable distribution P_z, with zero mean and variance ρ_z. The $\varphi^{(l)}$ are element-wise activation functions that can be either deterministic or stochastic. It will be useful to define the hidden variables $\mathbf{h}^{(l)} \in \mathbb{R}^{k_l}$ obtained from the output of layer $l-1$:

$$\mathbf{h}^{(l+1)} = \varphi^{(l)}\Big(\frac{1}{\sqrt{k_l}}\mathbf{W}^{(l)}\mathbf{h}^{(l)}\Big) \quad \Leftrightarrow \quad \mathbf{h}^{(l+1)} \sim P_{\text{out}}^{(l)}\Big(\ \cdot\ \Big|\frac{1}{\sqrt{k_l}}\mathbf{W}^{(l)}\mathbf{h}^{(l)}\Big).$$

We naturally let $\mathbf{h}^{(1)} \equiv \mathbf{z}$ and $\mathbf{h}^{(L+1)} \equiv \mathbf{v}$. We wrote in the last equation an equivalent parametrization in terms of channel densities $P_{\text{out}}^{(l)}$ over $\mathbb{R}^{k_{l+1}}$, that are also applied element-wise, and that parametrize the input/output relationship at each layer of the generative network. For instance, a deterministic layer l with non-linearity $\varphi^{(l)}$ is fully characterized by the scalar density $P_{\text{out}}^{(l)}(x|z) = \delta(x - \varphi^{(l)}(z))$.

Recall that the observer has access to the generative prior P_v, i.e. she/he knows the matrices $\mathbf{W}^{(l)} \in \mathbb{R}^{k_{l+1}\times k_l}$ and non-linearities $\varphi^{(l)}$. In the case of Model 5.2, she/he has similarly access to the details of P_u. The spike $\mathbf{v}^*$ is generated using a ground-truth value of the latent variable $\mathbf{z}^*$.

The spike is then estimated from the knowledge of the data matrix $\mathbf{Y}$, and the known form of the spiked-matrix model and of the generative prior. Only the spike $\mathbf{v}^*$ (and $\mathbf{u}^*$ for Model 5.2) and the latent vector $\mathbf{z}^*$ are unknown, and are to be inferred.

For concreteness and simplicity, the generative model that will be analyzed in the majority of examples given in this chapter is the single-layer case, i.e. eq. (5.3) with $L = 1$. In this case we define the *total compression ratio* of the generative prior as $\alpha \equiv p/k$. In what follows we will illustrate our results for φ being linear, sign and ReLU functions.

5.1.4 Summary of main results

Before diving into the details let us briefly summarize the main results of this chapter.

- We first conduct an asymptotic analysis for the performance of the optimal estimators (both information-theoretic and algorithmic) for the spiked-matrix Models 5.1 and 5.2. This analysis is based on a rigorous expression for the *mutual information* between the matrix $\mathbf{Y}$ and a general spike $\mathbf{v}^\star$ drawn from an arbitrary distribution P_v in $\mathbb{R}^p$. As in Chapter 4, this expression can be obtained by use of the replica method, but we will focus in this chapter on its rigorous derivation. Evaluating this expression on the generative priors discussed in Section 5.1.3, we obtain the threshold $\Delta_c \equiv \Delta_{\text{Algo}}$ below which the spike $\mathbf{v}^\star$ can be partially (or *weakly*) reconstructed in polynomial time.

- In order to understand the algorithmic limits of this problem, we derive an approximate message-passing (AMP) algorithm for Model 5.1 and 5.2. For all the generative architectures we analyze, we show that is attains the information-theoretic optimal performance.

- Finally, we propose a simple spectral method derived from our AMP algorithm, and we argue that it reaches the same weak-recovery threshold Δ_c. Remarkably, in certain cases we are able to rigorously analyze this spectral method independently of AMP using only random matrix theory tools.

Our main findings are in stark contrast to the known results for sparse PCA, and we therefore emphasize two important conclusions of our analysis:

(i) **No algorithmic gap with generative priors** – Sharp and detailed results are known in the thermodynamic limit (as defined above) when the spike $\mathbf{v}^\star \in \mathbb{R}^p$ is sampled from a *separable* distribution P_v. A detailed account of several examples can be found in [LKZ17]. The main finding for sparse priors P_v with k non-zero components is that when the sparsity $\rho = k/p = 1/\alpha$ is large enough then there exist optimal algorithms [DM14a], while for small enough ρ there is a striking gap between statistically optimal performance and the one of best known algorithms [LKZ15]. These conclusions are consistent with the well-known results for exact recovery of the support of $\mathbf{v}^\star$ [AW09a, BR13], which is one of the best-known cases in which gaps between statistical and best-known algorithmic performance were described.

Our analysis of the spiked-matrix models with generative priors reveals that in this setting, iterative algorithms are able to obtain asymptotically optimal performance even when the compression factor is important, i.e. $\alpha \gg 1$. It therefore suggests that generative priors are "better than sparsity" in the sense that they lead to algorithmically easier problems[2].

(ii) **Spectral algorithms reaching statistical threshold** – Arguably the most basic algorithm used to solve a spiked-matrix model is based on the leading singular vectors of

[2] Analogous conclusions about the lack of algorithmic gaps for the problem of phase retrieval under a generative prior can be found in [HLV18, ALB+20].

the matrix $\mathbf{Y}$. We will generically refer to this strategy as PCA. Previous works on spiked-matrix models [PWBM16, LKZ17] established that for separable priors, PCA reaches the optimal (among polynomial-time algorithms) weak-recovery threshold Δ_c, below which it is able to provide positive correlation between its estimator and the ground-truth spike. While for sparse priors positive correlation is statistically reachable for values $\Delta > \Delta_c$, no efficient algorithm beating the PCA threshold is known[3].

In the case of generative priors we find that well-chosen spectral methods can improve on canonical PCA. We design a spectral method, called LAMP, that reaches the statistically optimal threshold, meaning that for larger values of noise variance no other (even exponential) algorithm is able to reach positive correlation with the spike. This is another striking difference with sparse separable priors, making the generative priors algorithmically more attractive. We moreover demonstrate the performance of LAMP on real data, and show considerable improvement over canonical PCA, even though real data does not arise from a random generative prior.

Section 5.2 is dedicated to the analysis of information-theoretic and algorithmic optimal estimation, while Section 5.3 focuses on the spectral methods and our random matrix theory analysis. Section 5.4 is devoted to the proofs of the random matrix analysis.

5.2 Analysis of optimal estimation

5.2.1 Mutual information: the replica method rigorous once again

For conciseness, the following information-theoretic results are given for the Wigner model 5.1. They can be fully generalized to the Wishart case, and we refer the reader to [ALM+20] for more details on this regard.

From an optimization perspective, the problem we want to solve is to find the estimator $\mathbf{v}^\star$ that minimizes the mean squared error (MSE)

$$\mathrm{mse}(\Delta) \equiv \mathbb{E}||\hat{\mathbf{v}} - \mathbf{v}^\star||_2^2. \tag{5.4}$$

Since the information about the generative model P_v of the spike is given, we know from Section 1.1 that the estimator minimizing eq. (5.4) is given by the mean of the posterior distribution of the spike $\mathbb{P}(\mathbf{v}|\mathbf{Y})$, which we can write from Bayes' rule as

$$\mathbb{P}(\mathbf{v}|\mathbf{Y}) = \frac{1}{\mathbb{P}(Y)} P_v(\mathbf{v}) \prod_{1\le i<j\le p} \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{1}{2\Delta}\left(Y_{ij} - \frac{v_i v_j}{\sqrt{p}}\right)^2} \prod_{i=1}^{p} \frac{1}{\sqrt{4\pi\Delta}} e^{-\frac{1}{4\Delta}\left(Y_{ii} - \frac{v_i^2}{\sqrt{p}}\right)^2}. \tag{5.5}$$

The expression above is written in full generality, and for the time being we have not assumed anything about P_v. The naive approach of estimating $\hat{\mathbf{v}}^{\mathrm{opt}}$ from exact sampling of the posterior is intractable numerically, especially in the large-dimensional limit $p \to \infty$ of interest. However, it is still possible to track the performance of the optimal estimator without direct sampling, through the I-MMSE theorem connecting the *minimal mean squared error* (MMSE) to a derivative of the mutual information between the signal and the data [GSV05]. Note that this is completely equivalent to studying the free entropy in the statistical physics language.

Following this rationale, our first main result is a rigorous expression for the mutual information between the ground-truth spike $\mathbf{v}^\star$ and the observation $\mathbf{Y}$, valid in the thermodynamic limit $p \to \infty$. We state it informally, while a full technical statement can be found in [ALM+20]:

[3] This result holds for sparsity $\rho = \Theta(1)$. A line of work shows that when the sparsity k scales as $o(p)$, there exists algorithm that can outperform PCA in terms of weak recovery [AW09a, DM14b].

Theorem 5.1 (*Mutual information for Model 5.1 with structured spike, informal*)

Assume that the spike $\mathbf{v}^\star$ comes from a sequence (of growing dimension p) of structured priors P_v on $\mathbb{R}^p$. We define the *mutual information* as $I(\mathbf{Y};\mathbf{v}^\star) \equiv D_{\rm KL}(P_{(\mathbf{v}^\star,\mathbf{Y})}|P_{\mathbf{v}^\star}P_{\mathbf{Y}})$. Then

$$\lim_{p\to\infty} i_p \equiv \lim_{p\to\infty} \frac{I(Y;\mathbf{v}^\star)}{p} = \inf_{q_v\in(0,\rho_v)} i_{\rm RS}(\Delta, q_v),$$
$$\text{with} \quad i_{\rm RS}(\Delta, q_v) \equiv \frac{(\rho_v - q_v)^2}{4\Delta} + \lim_{p\to\infty} \frac{I(\mathbf{v};\mathbf{v}+\sqrt{\frac{\Delta}{q_v}}\mathbf{w})}{p}. \tag{5.6}$$

Here, $\mathbf{w}\sim\mathcal{N}(0,\mathrm{I}_p)$, and $\rho_v \equiv \lim_{p\to\infty} \mathbb{E}_{P_v}[\mathbf{v}^\intercal\mathbf{v}]/p$.

The proof for this theorem can be found in [ALM+20]. It uses interpolation techniques, but as opposed to the proof of the replica prediction in Chapter 4, interpolation is here not adaptive, and the arguments are therefore significantly simpler. The subscript of $i_{\rm RS}$ denotes replica symmetry, acknowledging that this prediction is first obtained by the replica method before being put on rigorous ground.

Let us first draw the consequences of Theorem 5.1. It connects the asymptotic mutual information of the spiked model with generative prior P_v to the mutual information between $\mathbf{v}$ taken from P_v and its noisy version, $I(\mathbf{v};\mathbf{v}+\sqrt{\Delta/q_v}\mathbf{w})$. As mentioned before, the mutual information is connected to the performance of the optimal estimator, and one can prove that for the spiked-matrix model (cf. [EAK18]) the MMSE on the spike $\mathbf{v}^\star$ is asymptotically given by:

$$\text{MMSE}_v = \rho_v - q_v^\star, \tag{5.7}$$

where $q_v^\star$ is the minimizer of $i_{\rm RS}(\Delta, q_v)$ that appears in Theorem 5.1. Computing $i_{\rm RS}$ is itself a high-dimensional task, hard in full generality, but it can be done for a range of non-trivial P_v:

- The simplest tractable case is when the prior P_v is separable, then it yields back exactly the previously known formula from [KXZ16, BDM+16, LM19].
- For the correlated Gaussian generative model $P_v = \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$, one can easily compute $I(\mathbf{v};\mathbf{v}+\sqrt{\Delta/q_v}\mathbf{w}) = \mathrm{Tr}[\ln(\mathrm{I}_p + q_v\mathbf{\Sigma}/\Delta)]/2$.
- More interestingly, let us consider the multi-layer generative prior with random weights from eq. (5.3). The single-layer formula (i.e. when $L=1$) for $i_{\rm RS}$ has been derived and proven in [BKM+19]. For the multi-layer $L\geq 1$ case the mutual information formula has been derived in [MKMZ17, Ree17] and proven for the case of two layers in [GML+19]. In [ALM+20], we showed that these previous works yield the following formula for the spiked Wigner Model 5.1 with multi-layer generative prior given by eq. (5.3):

$$i_{\rm RS}(\Delta, q_v) = \frac{\rho_v^2 + q_v^2}{4\Delta} + \frac{1}{\alpha}\operatorname*{extr}_{\{\hat{q}_l, q_l\}}\Big[\frac{1}{2}\sum_{l=1}^{L}\alpha_l\hat{q}_l q_l - \sum_{l=1}^{L}\alpha_{l+1}\Psi_{\rm out}^{(l)}(\hat{q}_{l+1}, q_l) - \Psi_z(\hat{q}_z)\Big]. \tag{5.8}$$

where $\alpha_l = k_l/k$ (in particular $\alpha_1 = 1$ and $\alpha_{L+1} = \alpha$). We also defined $\hat{q}_{L+1} \equiv q_v/\Delta$, $q_z = q_1$, $\hat{q}_z = \hat{q}_1$, and the functions $\Psi_z, \Psi_{\rm out}$ are defined by

$$\begin{cases} \Psi_z(x) & \equiv \mathbb{E}_\xi[\mathcal{Z}_z(x^{1/2}\xi, x)\ln(\mathcal{Z}_z(x^{1/2}\xi, x))], \\ \Psi_{\rm out}^{(l)}(x,y) & \equiv \mathbb{E}_{\xi,\eta}[\mathcal{Z}_{\rm out}^{(l)}(x^{1/2}\xi, x, y^{1/2}\eta, \rho_l - y)\times\ln(\mathcal{Z}_{\rm out}^{(l)}(x^{1/2}\xi, x, y^{1/2}\eta, \rho_l - y))], \end{cases}$$

with $\xi, \eta \sim \mathcal{N}(0,1)$, ρ_l is the second moment of the hidden variable h_l and $\mathcal{Z}_z$, $\mathcal{Z}_{\text{out}}^{(l)}$ are the normalizations of the following two denoising scalar distributions:

$$\begin{cases} Q_z(z;\gamma,\Lambda) & \equiv \dfrac{P_z(z)e^{-\frac{\Lambda}{2}z^2+\gamma z}}{\mathcal{Z}_z(\gamma,\Lambda)}, \\ Q_{\text{out}}^{(l)}(v,x;B,A,\omega,V) & \equiv \dfrac{P_{\text{out}}^{(l)}(v|x)e^{-\frac{A}{2}v^2+Bv}}{\mathcal{Z}_{\text{out}}^{(l)}(B,A,\omega,V)} \dfrac{e^{-\frac{(x-\omega)^2}{2V}}}{\sqrt{2\pi V}}. \end{cases} \tag{5.9}$$

5.2.2 Optimal performance and statistical thresholds: phase diagrams

As many results of the replica method, Theorem 5.1 combined with eq. (5.8) is remarkable in that it connects the asymptotic mutual information of a high-dimensional model with a simple scalar formula that can be easily evaluated. Moreover, it fully characterizes the statistical and algorithmic performance of the optimal estimators, allowing us to readily identify the thresholds separating the region between possible and impossible inference of the spike. Let us now draw some of its consequences for common activations. Taking the extremization over q_v and $(\hat{q}_l, q_l)_{1\le l\le L}$ in eq. (5.8), we obtain the following system of coupled equations:

$$\begin{cases} q_v = 2\partial_x \Psi_{\text{out}}^{(L)}(q_v/\Delta, q_L) \\ q_L = 2\partial_x \Psi_{\text{out}}^{(L-1)}(\hat{q}_L, q_{L-1}) \\ \quad\vdots \\ q_l = \partial_x \Psi_{\text{out}}^{(l-1)}(\hat{q}_l, q_{l-1}) \\ \quad\vdots \\ q_z = 2\Psi_z'(\hat{q}_z) \end{cases} \qquad \begin{cases} \hat{q}_L = 2\frac{\alpha}{\alpha_L}\partial_y \Psi_{\text{out}}^{(L)}(q_v/\Delta, q_L) \\ \hat{q}_{L-1} = 2\frac{\alpha_L}{\alpha_{L-1}}\tilde{\alpha}_{L-1}\partial_y \Psi_{\text{out}}^{(L-1)}(\hat{q}_L, q_{L-1}) \\ \quad\vdots \\ \hat{q}_l = 2\frac{\alpha_{l+1}}{\alpha_l}\partial_y \Psi_{\text{out}}^{(l)}(\hat{q}_{l+1}, q_l) \\ \quad\vdots \\ \hat{q}_z = 2\frac{\alpha_2}{\alpha_1}\partial_y \Psi_{\text{out}}^{(1)}(\hat{q}_2, q_z) \end{cases}, \tag{5.10}$$

As previously discussed, the set of solutions of these equations provides all the information about the performance of the Bayes-optimal estimator through eq. (5.7), and of the optimal algorithmic estimator as we will see. We shall refer to them (with a bit of anticipation) as the *state evolution* (SE) equations.

Weak-recovery threshold

An important first question that can be answered from eq. (5.10) is *weak recovery*, i.e. when is it possible to perform better than a random guess from the prior distribution P_v. For instance, we intuitively expect that when the prior is not biased towards a particular direction in $\mathbb{R}^p$ and for very high noise $\Delta \gg 1$ weak recovery is impossible. In terms of fixed points of eq. (5.10), this situation corresponds to the existence of the *non-informative* fixed point $q_v^\star = 0$ (i.e. maximum $\text{MSE}_v = \rho_v$, or zero overlap with the spike). Evaluating the right-hand side of eqs. (5.10) at $q_v = 0$, we can see that $q_v^\star = 0$ is a fixed point iff

$$\mathbb{E}_{P_z}[z] = 0 \quad \text{and} \quad \mathbb{E}_{Q_{\text{out}}^{(l),0}}[v] = 0, \tag{5.11}$$

where $Q_{\text{out}}^{(l),0}(v,x) \equiv Q_{\text{out}}^{(l)}(v,x;0,0,0,\rho_l)$ from eq. (5.9). Note that for multi-layer networks with deterministic channels and $\varphi^{(l)} \equiv \varphi$ for all l, the second condition is equivalent to φ being an odd function.

When the condition of eq. (5.11) holds, $(q_v, q_L, \hat{q}_L, \ldots, \hat{q}_z, q_z) = (0,0,0,\ldots,0,0)$ is a fixed point of eq. (5.10). In [ALM+20] we perform a detailed linear stability analysis of this fixed point. This gives a generic way to compute the weak-recovery threshold Δ_c. Note that as we use a

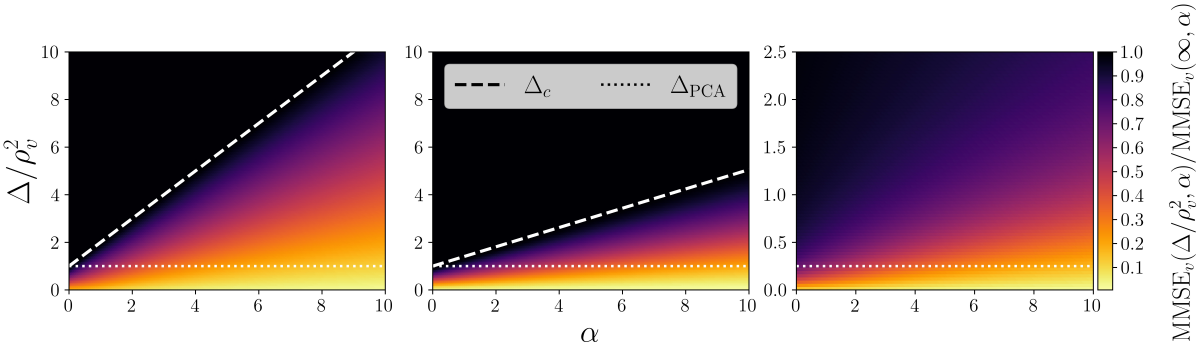


FIGURE 5.1: (Spiked Wigner model) MMSE_v as a function of noise to signal ratio Δ/ρ_v^2, and single-layer generative prior with compression ratio α for linear (left), sign (center), and ReLU (right) activations. Dashed white lines mark the transition Δ_c for odd activation. Dotted white lines mark the weak-recovery transition of canonical PCA.

local analysis of the uninformative fixed point, Δ_c corresponds to the *algorithmic* weak-recovery threshold: indeed, the performance of message-passing algorithms will be given by a local descent on i_{RS} starting from $q = 0$, as we will see in Section 5.2.3[4], while there might exists another global minimum of i_{RS} which does not arise from local instability of the $q = 0$ point. However, in all cases we investigated this situation did not occur, so that Δ_c corresponded both to the information-theoretic and the algorithmic weak-recovery threshold.

Single-layer generative prior

We first consider a single-layer generative prior $L = 1$. Fix $P_z = \mathcal{N}(0, 1)$ and $P_{\text{out}}^{(1)}(v|x) = \delta(v - \varphi(x))$, for $\varphi \in \{\text{linear}, \text{sign}, \text{ReLU}\}$. The first two choices of nonlinearity are odd, and therefore in these cases we expect a sharp weak-recovery transition as discussed above. This transition can be computed from the Jacobian (details of this computation are given in [ALM+20]) as we mentioned. We obtain $\Delta_c = 1 + \alpha$ for linear activation and $\Delta_c = 1 + 4\alpha/\pi^2$ for sign activation. In both cases, since $\alpha > 0$, it is clear that knowledge of the generative prior improves reconstruction, as the weak-recovery threshold of canonical PCA is $\Delta_{\text{PCA}} = 1$. Moreover, the larger α (i.e. the smaller the latent dimension with respect to the signal dimension), the better the reconstruction.

Fig. 5.1 summarizes this discussion. We numerically solve eq. (5.10), and plot the MMSE obtained from the solution in a heat map, for the linear, sign and ReLU activations. The white dashed line marks the threshold Δ_c obtained analytically as mentioned above. The property that we find the most striking is that in these three evaluated cases, for all values of Δ and α that we analyzed, we always found that eq. (5.10) has a unique stable solution. Thus we have not identified, in the physics terminology, any first order phase transition. Figure 5.2 shows examples of numerical MMSE curves for three activations discussed, and different values of α. The fixed point equations were solved iteratively from uncorrelated initial condition, and from initial condition corresponding to the ground truth signal, and we found that both lead to the same solution, illustrating the absence of computational gap.

Deeper generative priors

This observation generalizes to deeper $L > 1$ generative priors. Consider $P_z = \mathcal{N}(0, 1)$ and layer-wise constant activation $P_{\text{out}}^{(l)}(v|x) = \delta(v - \varphi(x))$. For the previous odd activation functions

[4]This is an important and generic property of AMP algorithms and their state evolution, that is also discussed in Chapters 4 and 6.

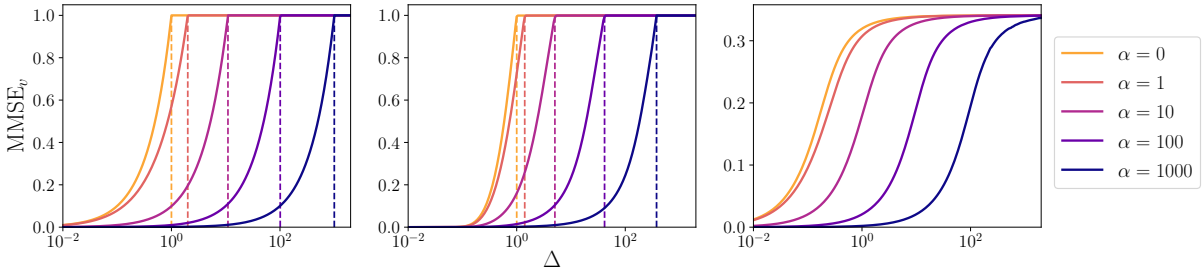


FIGURE 5.2: (Spiked Wigner model) MMSE_v as a function of noise Δ for $L = 1$ and a wide range of compression ratios $\alpha \in \{0, 1, 10, 100, 1000\}$, for linear (left), sign (center), and ReLU (right) activations. For the linear and sign functions, we show in dotted lines the weak-recovery transition Δ_c. In all cases we find a unique solution for eq. (5.10).

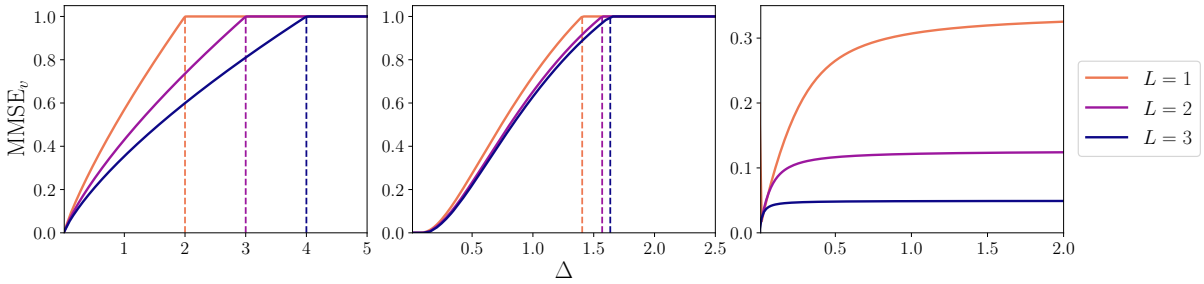


FIGURE 5.3: (Spiked Wigner model) MMSE_v as a function of noise Δ for $L = 1, 2, 3$ with constant compressive ratio $\alpha_1 = \alpha_2 = \alpha_3 = 1$, for linear (left), sign (center), and ReLU (right) activations. The second moment of v for $L = 1, 2, 3$ is $\rho_v^{(L)} = 1$ for linear and sign activations, while for ReLU $\rho_v^{(L)} = 1/2^L$. For linear and sign functions, we plot in dotted line the weak-recovery transition Δ_c.

discussed, we find that

- **Linear activation** – For $\varphi(x) = x$ the leading eigenvalue of the Jacobian becomes one at

$$\Delta_c \equiv 1 + \sum_{l=1}^{L} \frac{\alpha}{\alpha_l}. \tag{5.12}$$

  Note in particular that for $L = 1$ and in the limit $\alpha = 0$ we recover the transition of PCA $\Delta_c = 1$ known from the case with separable prior [LKZ17]. For $\alpha > 0$, we have $\Delta_c > 1$ meaning the spike can be estimated more efficiently when its structure is accounted for. In particular, the deeper the generative network for the spike, the easier estimation becomes.

- **Sign activation** – For $\varphi(x) = \text{sign}(x)$ the leading eigenvalue of the Jacobian becomes one at

$$\Delta_c = 1 + \sum_{l=1}^{L} \left(\frac{4}{\pi^2}\right)^l \frac{\alpha}{\alpha_l}. \tag{5.13}$$

  As in the linear case, we recover the threshold of canonical PCA for $\alpha = 0$, while for $\alpha > 0$ we can estimate the spike for larger noise values than in the separable case, and depth also improves estimation.

Fig. 5.3 illustrates our results for the multi-layer case. Note that we also didn't observe first order transitions for deeper networks, at least in the first-to-come-in-mind cases that we have investigated, i.e. deterministic deep networks with $\varphi^{(l)} \equiv \varphi \in \{\text{linear}, \text{sign}, \text{ReLU}\}$. However, we do not expect this behavior to be completely general neither. One can engineer a situation, for

Algorithm 4: AMP for the spiked Wishart model 5.2 with single-layer generative prior.

Result: The estimators $\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{z}}$

Input: Observation $\mathbf{Y} \in \mathbb{R}^{n\times p}$ and weight matrix $\mathbf{W} \in \mathbb{R}^{p\times k}$;

Initialize to zero: $(\mathbf{g}, \hat{\mathbf{u}}, \hat{\mathbf{v}}, \mathbf{B}_v, A_v, \mathbf{B}_u, A_u)^{t=0}$;

Initialize: $\hat{\mathbf{u}}^{t=1}, \hat{\mathbf{v}}^{t=1}, \hat{\mathbf{z}}^{t=1} \sim \mathcal{N}(0, \sigma^2)$, $\hat{\mathbf{c}}_u^{t=1} = \mathbf{1}_n$, $\hat{\mathbf{c}}_v^{t=1} = \mathbf{1}_p$, $\hat{\mathbf{c}}_z^{t=1} = \mathbf{1}_k$;

while *not converging* **do**

Spiked layer;

$\mathbf{B}_u^t = \frac{1}{\Delta}\frac{\mathbf{Y}}{\sqrt{p}}\hat{\mathbf{v}}^t - \frac{1}{\Delta}\frac{\mathbf{1}_p^\intercal \hat{\mathbf{c}}_v^t}{p}\mathrm{I}_n\hat{\mathbf{u}}^{t-1}$ and $A_u^t = \frac{1}{\Delta}\frac{\|\hat{\mathbf{v}}^t\|_2^2}{p}$;

$\mathbf{B}_v^t = \frac{1}{\Delta}\frac{\mathbf{Y}^\intercal}{\sqrt{p}}\hat{\mathbf{u}}^t - \frac{1}{\Delta}\frac{\mathbf{1}_n^\intercal \hat{\mathbf{c}}_u^t}{p}\mathrm{I}_p\hat{\mathbf{v}}^{t-1}$ and $A_v^t = \frac{1}{\Delta}\frac{\|\hat{\mathbf{u}}^t\|_2^2}{p}$;

Generative layer;

$V^t = \frac{1}{k}(\mathbf{1}_k^\intercal \hat{\mathbf{c}}_z^t)$ and $\boldsymbol{\omega}^t = \frac{1}{\sqrt{k}}\mathbf{W}\hat{\mathbf{z}}^t - V^t\mathbf{g}^{t-1}$ and $\mathbf{g}^t = f_{\text{out}}(\mathbf{B}_v^t, A_v^t, \boldsymbol{\omega}^t, V^t)$;

$\Lambda^t = \frac{1}{k}\|\mathbf{g}^t\|_2^2$ and $\boldsymbol{\gamma}^t = \frac{1}{\sqrt{k}}\mathbf{W}^\intercal\mathbf{g}^t + \Lambda^t\hat{\mathbf{z}}^t$;

Update of the estimated marginals;

$\hat{\mathbf{u}}^{t+1} = f_u(\mathbf{B}_u^t, A_u^t)$ and $\hat{\mathbf{c}}_u^{t+1} = \partial_B f_u(\mathbf{B}_u^t, A_u^t)$;

$\hat{\mathbf{v}}^{t+1} = f_v(\mathbf{B}_v^t, A_v^t, \boldsymbol{\omega}^t, V^t)$ and $\hat{\mathbf{c}}_v^{t+1} = \partial_B f_v(\mathbf{B}_v^t, A_v^t, \boldsymbol{\omega}^t, V^t)$;

$\hat{\mathbf{z}}^{t+1} = f_z(\boldsymbol{\gamma}^t, \Lambda^t)$ and $\hat{\mathbf{c}}_z^{t+1} = \partial_\gamma f_z(\boldsymbol{\gamma}^t, \Lambda^t)$;

$t = t + 1$;

end

instance with a very shifted ReLU on the last layer, and a very large intermediate layer, so that the spike $\mathbf{v}$ becomes effectively sparse with weakly correlated, almost independent, components, thus recovering the classical algorithmic gap [LKZ17].

So far we have only discussed the performance of the information theoretic optimal estimator - averting the question of estimating the signal itself. In the next section we close this gap by mean of an *approximate message-passing* (AMP) algorithm, again originating in our statistical physics toolbox.

5.2.3 Algorithmic optimal estimation

In this section we state and analyze an AMP algorithm tailored for spiked matrix estimation with generative priors. Its derivation from the belief propagation equations is fairly technical, and detailed in [ALM+20]. As we have discussed already in this thesis, AMP has two great virtues in this type of Bayes-optimal models: firstly, it achieves the best MSE among a wide class of general first order methods [CMW20], and secondly we can derive state evolution (SE) equations to track the MSE of AMP in the thermodynamic limit. As we will see, this MSE coincides with the optimal performance discussed in Section 5.2, even for large α. This result is particularly interesting when compared to the known performance of message-passing algorithms for sparse-PCA, for which AMP is not able to reach optimal statistical performance in the small sparsity regime [LML+17].

As underlined in [ALM+20], one can combine two AMPs for independent inference problems into a single one for a structured problem mixing the two. In particular, this applies to spiked-matrix estimation with single-layer generative prior, which can be seen as the combination of a *rank-one matrix factorization problem* [LKZ17] with a *generalized linear model* (GLM, cf. Section 1.1). Note that the multi-layer case can be derived by iterating this procedure, and the interested reader can refer to [ALM+20] for more details on this regard. We give the AMP algorithm for the spiked Wishart model 5.2 in Algorithm 4, with a single-layer generative prior.

We defined some auxiliary functions:

- f_u and f_z are the means of the distributions Q_u and Q_z, defined as

$$Q_u(u; B, A) \equiv \frac{P_u(u)e^{-\frac{1}{2}Au^2+Bu}}{\mathcal{Z}_u(B, A)} \qquad \text{and} \qquad Q_z(z; \gamma, \Lambda) \equiv \frac{P_z(z)e^{-\frac{\Lambda}{2}z^2+\gamma z}}{\mathcal{Z}_z(\gamma, \Lambda)}. \tag{5.14}$$

- f_v is the mean of v, and f_{out} is the mean of $V^{-1}(x-\omega)$, both with respect to

$$Q_{\text{out}}(v, x; B, A, \omega, V) \equiv \frac{e^{-\frac{1}{2}Av^2+Bv}P_{\text{out}}(v|x)}{\mathcal{Z}_{\text{out}}(B, A, \omega, V)}e^{-\frac{(x-\omega)^2}{2V}}. \tag{5.15}$$

The AMP algorithm for the spiked Wigner model 5.1 is very similar and can be readily obtained by imposing at each time step $(\hat{\mathbf{u}}^t, \hat{\mathbf{c}}_u^t) = (\hat{\mathbf{v}}^t, \hat{\mathbf{c}}_v^t)$ and removing the redundant equations in Algorithm 4.

5.2.4 State evolution equations

Perhaps the most important virtue of AMP-type algorithms is that their asymptotic performance (here the MSE) can be tracked exactly via a set of scalar equations called *state evolution* (SE). This has been proven for a range of models including the spiked matrix models with separable priors in [JM13], and with non-separable priors in [BMN20]. Adapting the steps of these works, we now derive the state evolution equations for our structured model. As for the algorithm, we state the equations for the Wishart model 5.2, from which the Bayes-optimal state evolution equations for the Wigner model 5.1 can be readily obtained.

We simply state the SE equations obtained by an asymptotic analysis of AMP. The SE gives the evolution of the following order parameters under Algorithm 4:

$$\begin{cases} q_u^t & \equiv \mathbb{E}_{\mathbf{u}^\star} \lim\limits_{n\to\infty} \dfrac{(\hat{\mathbf{u}}^t)^\intercal \hat{\mathbf{u}}^t}{n} = \mathbb{E}_{\mathbf{u}^\star} \lim\limits_{n\to\infty} \dfrac{(\hat{\mathbf{u}}^t)^\intercal \mathbf{u}^\star}{n} \equiv m_u^t, \\ q_v^t & \equiv \mathbb{E}_{\mathbf{v}^\star} \lim\limits_{p\to\infty} \dfrac{(\hat{\mathbf{v}}^t)^\intercal \hat{\mathbf{v}}^t}{p} = \mathbb{E}_{\mathbf{v}^\star} \lim\limits_{p\to\infty} \dfrac{(\hat{\mathbf{v}}^t)^\intercal \mathbf{v}^\star}{p} \equiv m_v^t, \\ q_z^t & \equiv \mathbb{E}_{\mathbf{z}^\star} \lim\limits_{k\to\infty} \dfrac{(\hat{\mathbf{z}}^t)^\intercal \hat{\mathbf{z}}^t}{k} = \mathbb{E}_{\mathbf{z}^\star} \lim\limits_{k\to\infty} \dfrac{(\hat{\mathbf{z}}^t)^\intercal \mathbf{z}^\star}{k} \equiv m_z^t. \end{cases} \tag{5.16}$$

Note that we used the Nishimori proposition 1.1, itself a consequence of Bayes-optimality, to deduce the equality between q and m. From this definition, it is clear that computing q_u, q_v, q_z is enough to access the asymptotic MSE achieved by the AMP algorithm. We can now state the closed set of AMP *state evolution equations* for the Wishart model 5.2:

$$\begin{cases} q_u^{t+1} = \mathbb{E}_{u^\star,\xi}\Big[f_u\Big(\dfrac{q_v^t}{\Delta}u^\star + \sqrt{\dfrac{q_v^t}{\Delta}}\xi, \dfrac{q_v^t}{\Delta}\Big)^2\Big], & \text{(5.17a)} \\ q_v^{t+1} = \mathbb{E}_{v^\star,\xi,\eta}\Big[f_v\Big(\dfrac{\tau q_u^t}{\Delta}v^\star + \sqrt{\dfrac{\tau q_u^t}{\Delta}}\xi, \dfrac{\tau q_u^t}{\Delta}, \sqrt{q_z^t}\eta, \rho_z - q_z^t\Big)^2\Big], & \text{(5.17b)} \\ \hat{q}_z^t = \alpha\mathbb{E}_{v^\star,\xi,\eta}\Big[f_{\text{out}}\Big(\dfrac{\tau q_u^t}{\Delta}v^\star + \sqrt{\dfrac{\tau q_u^t}{\Delta}}\xi, \dfrac{\tau q_u^t}{\Delta}, \sqrt{q_z^t}\eta, \rho_z - q_z^t\Big)^2\Big], & \text{(5.17c)} \\ q_z^{t+1} = \mathbb{E}_{z^\star,\xi}[f_z(\hat{q}_z^t z^\star + \sqrt{\hat{q}_z^t}\xi, \rho_z - q_z^t)^2]. & \text{(5.17d)} \end{cases}$$

Note that we introduced an additional auxiliary variable $\hat{q}_z$, and recall that $\tau \equiv n/p$, $\alpha = p/k$ and the definitions of the auxiliary functions in eqs. (5.14),(5.15). In the expectations, the variables ξ, η are taken independently from $\mathcal{N}(0, 1)$, and u^*, v^*, z^* are drawn from P_u, P_v, P_z. For a detailed derivation of eq. (5.17) the reader should refer to [ALM+20].

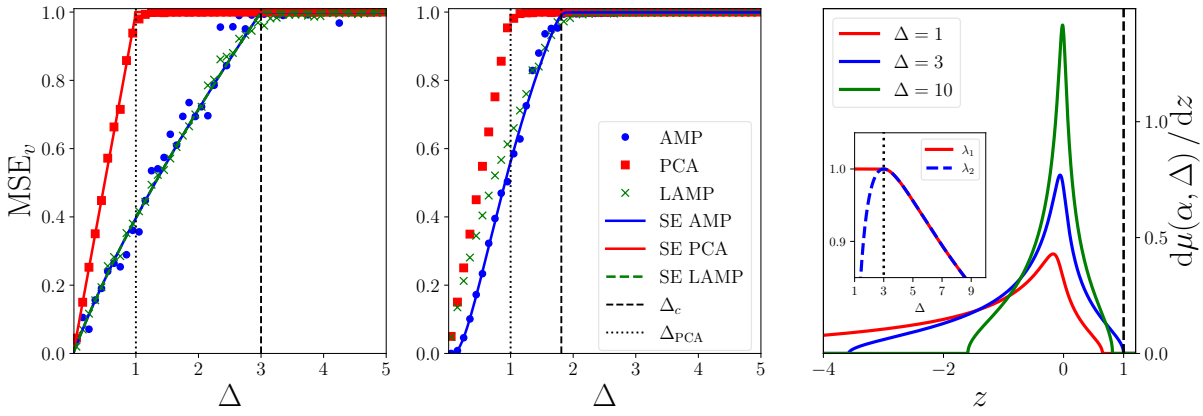


FIGURE 5.4: Comparison between PCA, LAMP and AMP for linear (left) and sign (center) activations, at compression ratio $\alpha = 2$. Full lines correspond to the asymptotic performances obtained from the SE equations, while dots correspond to simulations. We use $k = 10^4$ for LAMP, $k = 5.10^3$ for AMP. (Right) Illustration of the spectral phase transition in the matrix $\mathbf{\Gamma}_p^{vv}$ of eq. (5.21) at $\alpha = 2$ with informative leading eigenvector with eigenvalue equal to 1 out of the bulk for $\Delta \leq 1 + \alpha$. We show the bulk spectral density $\mu(\alpha, \Delta)$ and the two leading eigenvalues (inset).

State evolution equations in the Wigner model

The state evolution equations for the Wigner model 5.1 are obtained as a particular case of the above by simply restricting $q_u^t = q_v^t$ and $\tau = 1$. In the end (after performing a change of variables) this leaves us with only three coupled equations:

$$\begin{cases} q_z^{t+1} = \mathbb{E}_\xi[\{\mathcal{Z}_z \times f_z^2\}(\sqrt{\hat{q}_z^t}\xi, \hat{q}_z^t)] = 2\partial_{\hat{q}_z}\Psi_z(\hat{q}_z^t), & \text{(5.18a)} \\ \hat{q}_z^t = \alpha\mathbb{E}_{\xi,\eta}\Big[\{\mathcal{Z}_{\text{out}} \times f_{\text{out}}^2\}\Big(\sqrt{\frac{q_v^t}{\Delta}}\xi, \frac{q_v^t}{\Delta}, \sqrt{q_z^t}\eta, \rho_z - q_z^t\Big)\Big] = 2\alpha\partial_{q_z}\Psi_{\text{out}}\Big(\frac{q_v^t}{\Delta}, q_z^t\Big), & \text{(5.18b)} \\ q_v^{t+1} = \mathbb{E}_{\xi,\eta}\Big[\{\mathcal{Z}_{\text{out}} \times f_v^2\}\Big(\sqrt{\frac{q_v^t}{\Delta}}\xi, \frac{q_v^t}{\Delta}, \sqrt{q_z^t}\eta, \rho_z - q_z^t\Big)\Big] = 2\partial_{q_v}\Psi_{\text{out}}\Big(\frac{q_v^t}{\Delta}, q_z^t\Big). & \text{(5.18c)} \end{cases}$$

We initialize these iterations as $q_v^{t=0} = \varepsilon$, $q_z^{t=0} = \varepsilon$, with a small $\varepsilon > 0$[5]. We notice immediately that eq. (5.18) is identical to the fixed point equations related to the Bayes-optimal estimation, eq. (5.10) with specific time-indices and initialization, but crucially the same fixed points. Thus the analysis of fixed points in Section 5.2.2 applies straightforwardly here. In particular, since in all cases analyzed we found the stable fixed point of eq. (5.10) to be unique, we conclude that our AMP algorithm reaches asymptotically optimal performance in these cases.

We further check our results by numerically comparing the runs of AMP on finite size instances with the state evolution curves already presented in Fig. 5.2, also giving an idea of the amplitude of the finite size effects (which are found to be fairly small). This experiment is illustrated in Fig. 5.4, together with a curve for PCA and for LAMP, a spectral method we derive from AMP in the next section. A code for reproducing this experiment is provided in a GitHub repository [ALM+19].

[5]This is necessary as, by symmetry of the problem, the point $q_v = q_z = 0$ is always a fixed point of the SE.

5.3 LAMP: a spectral algorithm for generative priors

Spectral algorithms are arguably the most popular and simplest methods for solving the spiked matrix estimation problem. A seminal result from Baik, Ben Arous and Péché (BBP) [BBAP05] shows that the leading eigenvector of $\mathbf{Y}$ (i.e. PCA) is correlated with the signal if and only if the signal-to-noise ratio satisfies $\rho_v^2/\Delta > 1$. For sparse separable priors (with $\rho_v^2 = \Theta(1)$), $\Delta_{\rm PCA} = \rho_v^2$ is also the threshold for AMP and it is conjectured that no polynomial algorithm can improve upon it [LKZ17]. In contrast, in Section 5.2 we have developed an AMP algorithm (Algorithm 4) that has consistently better performance than PCA for structured priors, and in particular achieves the optimal weak recovery threshold.

Despite all its virtues, AMP is unarguably a convoluted algorithm. It would be desirable to design a simple spectral algorithm that still takes into account the structured nature of the prior. In this section we design such a spectral algorithm, hereafter named L-AMP, and we show that it matches the optimal weak-recovery threshold. Our derivation follows a strategy pioneered in [KMM+13], consisting on analyzing the AMP equations linearized around the non-informative fixed point. In this section, the discussion is framed on the Wigner model, the Wishart case being a straightforward generalization, given in [ALM+20]. We also focus on the single-layer generative prior, and we will describe how multi-layer generalizations can be made.

We discussed in Section 5.2 the existence condition of the uninformative fixed point in the Bayes-optimal SE equations. Not surprisingly, since the SE of AMP is identical to the Bayes-optimal one, the same conditions can be obtained independently from the AMP equations by analyzing when $\hat{\mathbf{v}} = \mathbf{0}$, and we find back eq. (5.11):

$$(\hat{\mathbf{v}}, \hat{\mathbf{z}}) = (\mathbf{0}, \mathbf{0}) \quad \text{is a fixed point of AMP iff} \quad \{\mathbb{E}_{Q^0_{\rm out}}[v] = 0 \quad \text{and} \quad \mathbb{E}_{P_v}[z] = 0\}. \tag{5.19}$$

5.3.1 Linearizing the AMP equations

The linearization of Algorithm 4 around the uninformative fixed point under eq. (5.19) is fairly straightforward, and the interested reader can find it in [ALM+20]. In the end, dropping time indices, we obtain an equation $\hat{\mathbf{v}} = \mathbf{\Gamma}_p^{vv}\hat{\mathbf{v}}$ where the *LAMP operator* $\mathbf{\Gamma}_p^{vv}$ is given by

$$\mathbf{\Gamma}_p^{vv} \equiv \frac{1}{\Delta}\Big((a-b)\mathrm{I}_p + b\frac{\mathbf{W}\mathbf{W}^\intercal}{k} + c\frac{\mathbf{1}_p\mathbf{1}_k^\intercal}{k}\frac{\mathbf{W}^\intercal}{\sqrt{k}}\Big)\Big(\frac{\mathbf{Y}}{\sqrt{p}} - a\mathrm{I}_p\Big), \tag{5.20}$$

where the parameters are simply the moments of distributions P_z and $Q^0_{\rm out}$

$$a \equiv \mathbb{E}_{Q^0_{\rm out}}[v^2] = \rho_v, \quad b \equiv \rho_z^{-1}\mathbb{E}_{Q^0_{\rm out}}[vx]^2, \quad c \equiv \frac{1}{2}\rho_z^{-3}\mathbb{E}_{P_z}[z^3]\mathbb{E}_{Q^0_{\rm out}}[vx^2]\mathbb{E}_{Q^0_{\rm out}}[vx].$$

Note that in most of the cases we studied, the parameter c, taking into account the skewness of the variable $\mathbf{z}$, is zero, simplifying considerably the structured matrix. Moreover, for the specific examples already discussed in Section 5.2, the LAMP operator $\mathbf{\Gamma}_p^{vv}$ is very simple. For instance, for Gaussian z and $P_{\rm out}(v|x) = \delta(v - \mathrm{sign}(x))$, we have $(a, b, c) = (1, 2/\pi, 0)$. Instead, for linear activation we get $(a, b, c) = (1, 1, 0)$. In this latter case, the LAMP operator can be written as

$$\mathbf{\Gamma}_p^{vv} = \frac{1}{\Delta}\mathbf{K}_p\,[\frac{\mathbf{Y}}{\sqrt{p}} - \mathrm{I}_p] \quad \text{with} \quad \mathbf{K}_p = \frac{[\mathbf{W}\mathbf{W}^\intercal]}{k} \simeq \mathbf{\Sigma} \equiv \frac{1}{n}\sum_\alpha \mathbf{v}^\alpha(\mathbf{v}^\alpha)^\intercal\,, \tag{5.21}$$

or, in other words, $\mathbf{K}_p$ is the covariance matrix of the structured spike $\mathbf{v}$. The same observation holds for the sign activation function. Interestingly, the covariance matrix $\mathbf{\Sigma}$ can be empirically estimated directly from samples of spikes, without the knowledge of the generative model (φ, W) itself, suggesting a simple practical implementation of LAMP. With this in mind, we use a more

Algorithm 5: LAMP (Linearized AMP) spectral algorithm

Input: Observed matrix $\mathbf{Y} \in \mathcal{S}_p$, prior P_v on $\mathbf{v} \in \mathbb{R}^p$;
Take the leading eigenvector $\hat{\mathbf{v}} \in \mathbb{R}^p$ of $\mathbf{\Gamma}_p^{vv} \equiv \mathbf{K}_p \left[\frac{\mathbf{Y}}{\sqrt{p}} - \mathrm{I}_p\right]$ with $\mathbf{K}_p = \mathbb{E}_{P_v}[\mathbf{v}\mathbf{v}^\intercal]$;

generic definition for LAMP as Algorithm 5. LAMP can therefore be interpreted as a refined version of PCA that takes into account the structure of the prior by incorporating the non-trivial correlations through $\mathbf{K}_p$ into the spectral estimation. In particular taking $P_v = \mathcal{N}(\mathbf{0}, \mathrm{I}_p)$, we obtain $\mathbf{\Gamma}_p^{vv} = [\mathbf{Y}/\sqrt{p} - \mathrm{I}_p]/\Delta$ and recognize the vanilla PCA operator (up to a shift).

State evolution for LAMP in the linear case – Analogously to the state evolution for AMP, the asymptotic performance of both PCA and LAMP can be evaluated in a closed form for the spiked Wigner model with single-layer generative prior with linear activation. The corresponding expressions are derived in Appendix D.4 and plotted in the left panel of Fig. 5.4.

A note on normalization – Note that the spectral methods and AMP use different normalizations. In order to compare them fairly in Fig. 5.4, we renormalized the spectral estimators to match the AMP estimator norm. We detail the reasons behind this renormalization in Appendix D.4.

5.3.2 Random matrix perspective on the spectral methods

Remarkably, the performance of the spectral method described in Algorithm 5 can be investigated independently of AMP using solely random matrix theory, for a linear generative prior. An analysis of the random matrix of eq. (5.21) shows that a spectral phase transition takes place as $\Delta_c = 1 + \alpha$ (as for AMP). This transition is analogous to the well-known BBP transition [BBAP05], but for the non-GOE random matrix of eq. (5.21).

For the spiked Wigner models with linear generative prior we prove two detailed theorems describing the behavior of the supremum of the bulk spectral density, the transition of the largest eigenvalue and the correlation of the corresponding eigenvector. The interested reader can find natural extensions of these theorems for spiked Wishart models in our work [ALM+20]. We assume in the following that $\rho_v = 1$ to simplify the analysis (without any loss of generality). Recall that we have $\alpha = p/k$ and

$$\mathbf{\Gamma}_p^{vv} \equiv \left[\frac{1}{k}\mathbf{W}\mathbf{W}^\intercal\right]\left[\frac{1}{\sqrt{\Delta p}}\boldsymbol{\xi} + \frac{1}{\Delta}\frac{\mathbf{v}\mathbf{v}^\intercal}{p} - \frac{1}{\Delta}\mathrm{I}_p\right]. \tag{5.22}$$

Here $\xi_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1 + \delta_{ij})$. Note that $\mathbf{v}$ also depends on $\mathbf{W}$ via the generative prior, so that the spike in eq. (5.22) is correlated with the unspiked matrix, as opposed to the classical case of [BBAP05]. We first describe the behavior of the bulk of $\mathbf{\Gamma}_p^{vv}$.

Theorem 5.2 (*Bulk of the spectral density of* $\mathbf{\Gamma}_p^{vv}$)

For any $\alpha, \Delta > 0$, as $p \to +\infty$, the spectral measure of $\mathbf{\Gamma}_p^{vv}$ converges almost surely and in the weak sense to a well-defined and compactly supported probability measure $\mu(\alpha, \Delta)$, and we denote supp μ its support. We separate two cases:

(*i*) If $\Delta \leq 1/4$, then supp $\mu \subseteq \mathbb{R}_-$.

(*ii*) Assume now $\Delta > 1/4$ and denote $z_1(\Delta) \equiv -\Delta^{-1} + 2\Delta^{-1/2} > 0$. Let ρ_Δ be the probability measure on $\mathbb{R}$ with density

$$\rho_\Delta(\mathrm{d}t) = \frac{\sqrt{\Delta}}{2\pi}\sqrt{4 - \Delta\Big(t + \frac{1}{\Delta}\Big)^2} \times \mathbb{1}\Big\{|t + \frac{1}{\Delta}| \leq \frac{2}{\sqrt{\Delta}}\Big\}\,\mathrm{d}t. \tag{5.23}$$

The following equation admits a unique solution for $s \in (-z_1(\Delta)^{-1}, 0)$:

$$\alpha \int \rho_\Delta(\mathrm{d}t)\Big(\frac{st}{1+st}\Big)^2 = 1. \tag{5.24}$$

We denote this solution as $s_{\mathrm{edge}}(\alpha, \Delta)$ (or simply s_{edge}). The supremum of the support of $\mu(\alpha, \Delta)$ is denoted $\lambda_{\max}(\alpha, \Delta)$ (or simply $\lambda_{\max}$). It is given by:

$$\lambda_{\max} = \begin{cases} \alpha \int \dfrac{\rho_\Delta(\mathrm{d}t)t}{1 + s_{\mathrm{edge}}t} - \dfrac{1}{s_{\mathrm{edge}}} & \text{if } \alpha \leq 1, \\ \max\Big(0, \alpha \int \dfrac{\rho_\Delta(\mathrm{d}t)t}{1 + s_{\mathrm{edge}}t} - \dfrac{1}{s_{\mathrm{edge}}}\Big) & \text{if } \alpha > 1. \end{cases}$$

As a function of Δ, $\lambda_{\max}$ has a unique global maximum, reached exactly at the point $\Delta = \Delta_c(\alpha) = 1 + \alpha$. Moreover, $\lambda_{\max}(\alpha, \Delta_c(\alpha)) = 1$.

We turn now to the description of the transition in the spectrum:

Theorem 5.3 (*Transition of the largest eigenvalue of* $\mathbf{\Gamma}_p^{vv}$)

Let $\alpha > 0$. We denote $\lambda_1 \geq \lambda_2$ the first and second eigenvalues of $\mathbf{\Gamma}_p^{vv}$.

- If $\Delta \geq \Delta_c(\alpha)$, then as $p \to \infty$ we have a.s. $\lambda_1 \to \lambda_{\max}$ and $\lambda_2 \to \lambda_{\max}$.
- If $\Delta \leq \Delta_c(\alpha)$, then as $p \to \infty$ we have a.s. $\lambda_1 \to 1$ and $\lambda_2 \to \lambda_{\max}$.

Further, denoting $\tilde{\mathbf{v}}$ a normalized ($\|\tilde{\mathbf{v}}\|^2 = p$) eigenvector of $\mathbf{\Gamma}_p^{vv}$ with eigenvalue λ_1, then a.s. $|\tilde{\mathbf{v}}^\intercal \mathbf{v}^\star|^2/p^2 \to \epsilon(\Delta)$ a.s., where $\epsilon(\Delta) = 0$ for all $\Delta \geq \Delta_c(\alpha)$, $\epsilon(\Delta) > 0$ for all $\Delta < \Delta_c(\alpha)$ and $\lim_{\Delta \downarrow 0} \epsilon(\Delta) = 1$.

Theorems 5.2 and 5.3 are illustrated in the right panel of Fig. 5.4.

As we will see, the proof gives the value of $\epsilon(\Delta)$, which coherently leads to the MSE of Fig. 5.4 in the linear case. The proofs of Theorems 5.2 and 5.3 are given in Sections 5.4.1 and 5.4.2. The method of proof of Theorem 5.3 is inspired by [BGN11][6], and allows us to compute the squared

[6] Note that while all the calculations are justified, refinements would be needed in order to be completely rigorous. These refinements would follow exactly some proofs of [SB95] and [BGN11], so we will refer precisely to them when necessary.

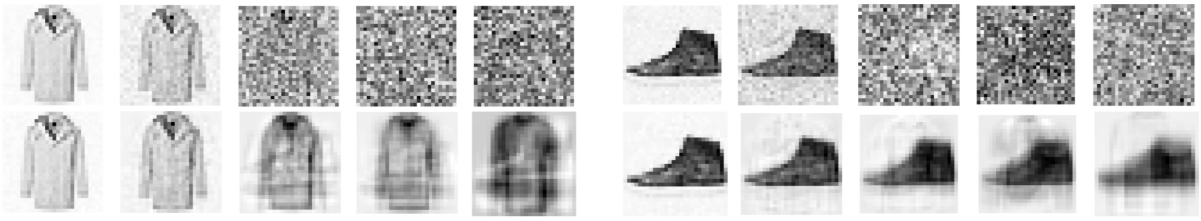


FIGURE 5.5: Illustration of canonical PCA (top line) and LAMP (bottom line) spectral methods on the spiked Wigner model. The covariance $\mathbf{\Sigma}$ is estimated empirically from the FashionMNIST database. The estimation of the spike is shown for two images from FashionMNIST, with (from left to right), noise variance $\Delta = 0.01, 0.1, 1, 2, 10$.

correlation $\epsilon(\Delta)$. It is given, for all $\Delta < \Delta_c(\alpha)$, as

$$\epsilon(\Delta) = \frac{1}{\alpha} \frac{[S^{(2)}(1)]^2}{S^{(1,2)}(1)}.$$

The functions $S^{(1,2)}$ and $S^{(2)}$ are defined in the following Lemma 5.5, and formulas are also given that allow to compute them numerically. A Mathematica demonstration notebook is provided in a GitHub repository [ALM+19].

Non-linear activation – In the non-linear case the random matrix analysis is harder to perform. Indeed, in the matrix $\mathbf{\Gamma}_p^{vv}$, the Wishart matrix $\mathbf{W}\mathbf{W}^\intercal/k$ is replaced by $a\mathrm{I} + b\mathbf{W}\mathbf{W}^\intercal/k$ with $a, b \geq 0$. It is thus not possible to relate the spectrum of $\mathbf{\Gamma}_p^{vv}$ to the one of a symmetric matrix of the type $\mathbf{W}\mathbf{Z}\mathbf{W}^\intercal$ with $\mathbf{W}$ a gaussian i.i.d. matrix (a *generalized sample covariance matrix* as defined in Section 1.5). Techniques from free probability allow for a detailed treatment of the spectral transition in this case, but such an analysis is not performed in this thesis.

5.3.3 Application to real data recovery

Remarkably, the LAMP operator in eq. (5.21) only depends on the generative prior through its covariance. An interesting exercise is therefore to apply LAMP to recover real data by simply using the empirical covariance for n samples of the spikes $\{\mathbf{v}^\alpha\}_{\alpha=1}^n$.

As an illustration we perform the following experiment: the spikes $\mathbf{v}^\star$ are drawn from the Fashion-MNIST dataset [XRV17], and are used to generate the spiked matrix $\mathbf{Y}$ according to eq. (5.1). We then apply LAMP (Algorithm 5) to reconstruct the spikes, repeating this experiment for different noise values Δ. In Fig. 5.5 we compare the reconstruction by LAMP with standard PCA over $\mathbf{Y}$. In principle, we have no theoretical guarantees about the performance of LAMP, since the Fashion-MNIST images are not drawn from the generative class studied above. Nevertheless, it is striking to observe that LAMP greatly outperforms PCA. A demonstration notebook illustrating this experiment is provided in a GitHub repository [ALM+19].

5.4 Random matrix analysis of the transition

In this section we present the proofs of Theorems 5.2 and 5.3.

5.4.1 The bulk of eigenvalues: proof of Theorem 5.2

Proof of Theorem 5.2 (*ii*) – We begin by treating the more involved case (*ii*), that is we assume $\Delta > 1/4$. Note first that by basic linear algebra, the spectrum of $\mathbf{\Gamma}_p^{vv}$ is, up to 0

eigenvalues, the same as the spectrum of the following matrix $\mathbf{\Gamma}_k^{vv}$:

$$\mathbf{\Gamma}_k^{vv} \equiv \frac{1}{k}\mathbf{W}^\intercal[\frac{1}{\sqrt{\Delta p}}\boldsymbol{\xi} + \frac{1}{\Delta}\frac{\mathbf{v}\mathbf{v}^\intercal}{p} - \frac{1}{\Delta}\mathrm{I}_p]\mathbf{W} \in \mathcal{S}_k. \tag{5.25}$$

More precisely, if $p \geq k$ we have $\mathrm{Sp}\,(\mathbf{\Gamma}_p^{vv}) = \mathrm{Sp}\,(\mathbf{\Gamma}_k^{vv}) \cup \{0\}^{p-k}$, and conversely if $k > p$. These additional zero eigenvalues in the case $\alpha > 1$ explain the $\max(0, \cdot)$ term in the expression of $\lambda_{\max}$ in Theorem 5.2.

For the remainder of the proof we consider $\mathbf{\Gamma}_k^{vv}$ instead of $\mathbf{\Gamma}_p^{vv}$ given the remark above. Moreover, for simplicity we will drop the vv exponent in those matrices, and just denote them $\mathbf{\Gamma}_k, \mathbf{\Gamma}_p$.

The bulk of $\mathbf{\Gamma}_k$ can be studied using standard random matrix theory results. Indeed, by the celebrated results of Wigner [Wig55], the spectral distribution of the matrix $\boldsymbol{\xi}/\sqrt{\Delta p} - \mathrm{I}_p/\Delta$ converges in law (and almost surely) as $p \to \infty$ to ρ_Δ, given by eq. (5.23). Moreover, $\mathbf{\Gamma}_k$ is precisely a *generalized sample covariance matrix* that we introduced in Section 1.5. Let us define $\nu(\alpha, \Delta)$ as the LSD of $\mathbf{\Gamma}_k$, and $g_\nu(z)$ its Stieltjes transform. The main quantity of interest to us is z_{edge}, defined as the supremum of the support of $\nu(\alpha, \Delta)$. It is easy to see that g_ν induces a strictly increasing diffeomorphism $g_\nu : (z_{\text{edge}}, +\infty) \to (\lim_{z \downarrow z_{\text{edge}}} g_\nu(z), 0)$, so that we can define its inverse g_ν^{-1} and from Theorem 1.7, it satisfies for every $s \in (\lim_{z \downarrow z_{\text{edge}}} g_\mu(z), 0)$:

$$g_\nu^{-1}(s) = -\frac{1}{s} + \alpha \int \rho_\Delta(\mathrm{d}t)\frac{t}{1+st}. \tag{5.26}$$

From the remarks above, $\mu(\alpha, \Delta)$ and $\nu(\alpha, \Delta)$ only differ by the addition of a delta distribution. If $z_{\text{edge}} \geq 0$, then it will also be the supremum of the support of $\mu(\alpha, \Delta)$, and thus equal to $\lambda_{\max}$.

In order to compute z_{edge} from eq. (5.26), we use a result of Section 4 of [SC95], also stated for instance in [LS16], that describes the form of the support of $\nu(\alpha, \Delta)$. It can be stated in the following way. Recall that since $\Delta > 1/4$, $z_1(\Delta) > 0$ is the maximum of the support of ρ_Δ. Let s_{edge} be the unique solution in $(-z_1(\Delta)^{-1}, 0)$ of the equation $(g_\nu^{-1})'(s) = 0$, that is by eq. (5.26):

$$\alpha \int \rho_\Delta(\mathrm{d}t)\Big(\frac{st}{1+st}\Big)^2 = 1. \tag{5.27}$$

Indeed, it is straightforward to show that the left-hand side of eq. (5.27) tends to 0 as $s \uparrow 0$, to $+\infty$ as $s \downarrow -z_1(\Delta)^{-1}$, and is a strictly decreasing and continuous function of s. Then (see for instance eqs. (2.13) and (2.14) of [LS16]) z_{edge} is given by

$$z_{\text{edge}} = \lim_{s \downarrow s_{\text{edge}}} g_\nu^{-1}(s) = -\frac{1}{s_{\text{edge}}} + \alpha \int \rho_\Delta(\mathrm{d}t)\frac{t}{1+s_{\text{edge}}t}.$$

This ends the proof of (ii). □

Let us make a final remark that will be useful in our future analysis. Note that $z_1(\Delta) > 1$ for all $\Delta > 1$. Moreover, for all $\Delta > 1$, we have by an explicit computation:

$$\alpha \int \rho_\Delta(\mathrm{d}t)\Big(\frac{t}{1-t}\Big)^2 = \frac{\alpha}{\Delta - 1}.$$

By the argument above, this yields the following result, that we state as a lemma:

Lemma 5.4

Assume $\Delta > 1$. If $\Delta < \Delta_c(\alpha)$, then $s_{\text{edge}} > -1$. Conversely, if $\Delta > \Delta_c(\alpha)$, then $s_{\text{edge}} < -1$. And moreover for $\Delta = \Delta_c(\alpha)$, $s_{\text{edge}} = -1$.

Proof of Theorem 5.2 (i) – Assume now $\Delta \leq 1/4$, so that $\mathrm{supp}(\rho_\Delta) \subset \mathbb{R}_-$. Since $0 \in \mathbb{R}_-$, we can use again the remark we made in the proof of (ii) to study $\mathbf{\Gamma}_k$ instead of $\mathbf{\Gamma}_p$. Moreover, Theorem 1.7 still applies here so that we have the Marchenko-Pastur equation (5.26). By the Stieltjes-Perron inversion formula (Theorem 1.5) it is enough to check that for every $z > 0$, there exists a unique $s < 0$ such that $g_\nu^{-1}(s) = z$. Indeed, this will yield that $s = g_\nu(z) \in \mathbb{R}$, so that $\lim_{\epsilon \downarrow 0} \mathrm{Im}\, g_\nu(z + i\epsilon) = 0$. As this holds for every $z > 0$, it implies $\mathrm{supp}(\nu) \subseteq \mathbb{R}_-$ and thus $\mathrm{supp}(\mu) \subseteq \mathbb{R}_-$. From eq. (5.26) and the fact that $\mathrm{supp}(\rho_\Delta) \subseteq \mathbb{R}_-$, we easily obtain

$$\lim_{s \to -\infty} g_\nu^{-1}(s) = 0, \quad \text{and} \quad \lim_{s \to 0^-} g_\nu^{-1}(s) = +\infty.$$

Moreover, $g_\nu^{-1}(s)$ is a strictly increasing continuous function of s, so that the existence and uniqueness of $s = g_\nu(z) < 0$ for any $z > 0$ is immediate, which ends the proof. □

We now prove the final statement of Theorem 5.2, on the behavior of $\lambda_{\max}$ with Δ, at fixed α. **Proof of the behavior of $\lambda_{\max}$ with Δ** – Recall that $z_{\mathrm{edge}} = -(1/s_{\mathrm{edge}}) + \alpha \int \rho_\Delta(\mathrm{d}t) t/(1 + s_{\mathrm{edge}} t)$. Then we know that $\lambda_{\max} = z_{\mathrm{edge}}$ if $\alpha \leq 1$ and $\lambda_{\max} = \max(0, z_{\mathrm{edge}})$ if $\alpha > 1$. Let us make a few remarks:

- We already showed that, if $\Delta \leq 1/4$ then $z_{\mathrm{edge}} \leq 0$.
- It is trivial by the form of $\mathbf{\Gamma}_k$ that, as $\Delta \to +\infty$, $z_{\mathrm{edge}} \to 0$.

It is easy to see that z_{edge} is a continuous and differentiable function of Δ, so that if we show the two following facts for any $\Delta \geq 1/4$:

$$\frac{\mathrm{d}z_{\mathrm{edge}}}{\mathrm{d}\Delta} = 0 \Leftrightarrow \Delta = \Delta_c(\alpha) = 1 + \alpha, \tag{5.28}$$

$$z_{\mathrm{edge}}(\Delta_c(\alpha)) = 1, \tag{5.29}$$

this would end the proof as z_{edge} would then have a unique global maximum, located in $\Delta = \Delta_c(\alpha)$, in which we have $\lambda_{\max} = 1$. We thus prove eq. (5.28) and eq. (5.29) in the following. □

Proof of eq. (5.28) – By the chain rule:

$$\frac{\mathrm{d}z_{\mathrm{edge}}}{\mathrm{d}\Delta} = \frac{\partial z_{\mathrm{edge}}}{\partial \Delta} + \frac{\partial s_{\mathrm{edge}}}{\partial \Delta} \frac{\partial z_{\mathrm{edge}}}{\partial s_{\mathrm{edge}}} = \frac{\partial z_{\mathrm{edge}}}{\partial \Delta}.$$

Indeed one can check $\partial z_{\mathrm{edge}}/\partial s_{\mathrm{edge}} = 0$ from eq. (5.27) and the Marchenko-Pastur equation. Given the explicit form of ρ_Δ, one can compute easily:

$$\frac{\partial z_{\mathrm{edge}}}{\partial \Delta} = -\frac{\alpha}{2 s_{\mathrm{edge}}^3} \Big[\frac{s_{\mathrm{edge}} + 2 s_{\mathrm{edge}}^2 - \Delta}{\sqrt{s_{\mathrm{edge}}^2 - 2 s_{\mathrm{edge}}(1 + 2 s_{\mathrm{edge}})\Delta + \Delta^2}} + 1 \Big].$$

It is then simple analysis to see that since $s_{\mathrm{edge}} < 0$, $\partial z_{\mathrm{edge}}/\partial \Delta = 0$ iff $s_{\mathrm{edge}} = -1$ and $\Delta > 1$. By Lemma 5.4, this is equivalent to $\Delta = \Delta_c(\alpha) = 1 + \alpha$. □

Proof of eq. (5.29) – By Lemma 5.4, we know that for $\Delta = \Delta_c(\alpha)$ we have $s_{\mathrm{edge}} = -1$. Given the definition of ρ_Δ in eq. (5.23), it is straightforward to compute:

$$z_{\mathrm{edge}}(\Delta_c(\alpha)) = -1 + \alpha \int \rho_{\Delta_c(\alpha)}(\mathrm{d}t) \frac{t}{1 - t} = 1.$$

□

5.4.2 BBP-like transition: proof of Theorem 5.3

Transition of the largest eigenvalue and corresponding eigenvector – This part is a detailed outline of the proof. Some parts of the calculation are not fully rigorous, however they can be justified more precisely by following exactly the lines of [BGN11] and [SC95]. We will emphasize when such refinements have to be made. Recall that we have by eq. (5.22) the following decomposition of $\mathbf{\Gamma}_p^{vv}$ (denoted $\mathbf{\Gamma}_p$ for simplicity):

$$\mathbf{\Gamma}_p = \underbrace{\Big[\frac{1}{k}\mathbf{W}\mathbf{W}^\intercal\Big]\Big[\frac{1}{\sqrt{\Delta p}}\boldsymbol{\xi} - \frac{1}{\Delta}\mathrm{I}_p\Big]}_{\mathbf{\Gamma}_p^{(0)}} + \underbrace{\frac{1}{\Delta}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\frac{\mathbf{v}\mathbf{v}^\intercal}{p}}_{\text{rank-1 perturbation}}. \tag{5.30}$$

Theorem 5.2, along with its proof, already describes in great detail the LSD of $\mathbf{\Gamma}_p^{(0)}$. Note that for any $\lambda \in \mathbb{R}$ that is not an eigenvalue of $\mathbf{\Gamma}_p^{(0)}$ one can write:

$$\det(\lambda \mathrm{I}_p - \mathbf{\Gamma}_p) = \det\Big(\mathrm{I}_p - (\lambda \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\frac{1}{\Delta}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\frac{\mathbf{v}\mathbf{v}^\intercal}{p}\Big)\det(\lambda \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)}).$$

In particular, this implies immediately that λ is an eigenvalue of $\mathbf{\Gamma}_p$ and not an eigenvalue of $\mathbf{\Gamma}_p^{(0)}$ if and only if 1 is an eigenvalue of $(\lambda \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\mathbf{W}\mathbf{W}^\intercal\mathbf{v}\mathbf{v}^\intercal/(\Delta k p)$. Since this is a rank-one matrix, its only non-zero eigenvalue is equal to its trace, so it is equivalent to:

$$1 = \mathrm{Tr}\,[(\lambda \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\frac{1}{\Delta}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\frac{\mathbf{v}\mathbf{v}^\intercal}{p}]. \tag{5.31}$$

Recall that by definition, $\mathbf{v}$ is constructed as $\mathbf{v} = \mathbf{W}\mathbf{z}/\sqrt{k}$, with $\mathbf{z} \sim \mathcal{N}(0, \mathrm{I}_k)$. For any matrix $\mathbf{A}$, we have the classical concentration $\mathbf{z}^\intercal\mathbf{A}\mathbf{z}/k = \mathrm{Tr}\mathbf{A}/k$ with high probability as $k \to \infty$. In eq. (5.31), this yields at leading order as $p \to \infty$:

$$\Delta = \frac{1}{p}\mathrm{Tr}\Big[(\lambda \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\Big(\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\Big)^2\Big]. \tag{5.32}$$

For practicality, we will prefer to use $k \times k$ matrices. We use the simple linear algebra identity, for any $p \times p$ symmetric matrix $\mathbf{A}$, and any integer $q \geq 1$:

$$\mathrm{Tr}\Big[\Big(\lambda \mathrm{I}_p - \frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{A}\Big)^{-1}\Big(\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\Big)^q\Big] = \mathrm{Tr}\Big[\Big(\lambda \mathrm{I}_k - \frac{1}{k}\mathbf{W}^\intercal\mathbf{A}\mathbf{W}\Big)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^q\Big].$$

This can be derived for instance by expanding both sides in powers of λ^{-1} and using the cyclicity of the trace. Using this along with eq. (5.32) we can state that the eigenvalues of $\mathbf{\Gamma}_p$ that are outside of the spectrum of $\mathbf{\Gamma}_p^{(0)}$ must satisfy, as $k \to \infty$:

$$\alpha\Delta = \frac{1}{k}\mathrm{Tr}\Big[(\lambda \mathrm{I}_k - \mathbf{\Gamma}_k^{(0)})^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^2\Big], \tag{5.33}$$

with

$$\mathbf{\Gamma}_k^{(0)} \equiv \frac{1}{k}\mathbf{W}^\intercal\Big[\frac{1}{\sqrt{\Delta p}}\boldsymbol{\xi} - \frac{1}{\Delta}\mathrm{I}_p\Big]\mathbf{W}.$$

We will now make use of two important lemmas, at the core of our analysis. They will also prove to be useful in the eigenvector correlation analysis.

Lemma 5.5 (*Hierarchy of $S_k^{(r)}$ functions*)

Recall that ν is the limit eigenvalue distribution of $\mathbf{\Gamma}_k^{(0)}$, that the supremum of its support is $\lambda_{\max}$, and its Stieltjes transform is g_ν. For every integer $r \geq 0$, we define:

$$S_k^{(r)}(\lambda) \equiv \frac{1}{k}\mathrm{Tr}\left[(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^r\right].$$

For $r \in \{0,1,2,3\}$[7] and every $\lambda > \lambda_{\max}$, as $k \to \infty$, $S_k^{(r)}(\lambda)$ converges almost surely to a well defined limit $S^{(r)}(\lambda)$. This limit is given by:

$$\begin{cases} S^{(0)}(\lambda) &= g_\nu(\lambda), \\ S^{(1)}(\lambda) &= g_\nu(\lambda)[\alpha - (1+\lambda g_\nu(\lambda))], \\ S^{(2)}(\lambda) &= g_\nu(\lambda)[\alpha(1+\alpha) - (1+2\alpha)(1+\lambda g_\nu(\lambda)) + (1+\lambda g_\nu(\lambda))^2], \\ S^{(3)}(\lambda) &= g_\nu(\lambda)[(\alpha + 3\alpha^2 + \alpha^3) + (2+3\alpha)(1+\lambda g_\nu(\lambda))^2 \\ &\qquad -(1+5\alpha+3\alpha^2)(1+\lambda g_\nu(\lambda)) - (1+\lambda g_\nu(\lambda))^3]. \end{cases}$$

We define similarly for every integers $r, q \geq 0$:

$$S_k^{(r,q)}(\lambda) \equiv \frac{1}{k}\mathrm{Tr}\left[(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^r(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^q\right].$$

Note that $S_k^{(r,q)} = S_k^{(q,r)}$ and that $S_k^{(r,0)}(\lambda) = \partial_z S_k^{(r)}(\lambda)$. For every $\lambda > \lambda_{\max}$, $S_k^{(1,1)}(\lambda)$ and $S_k^{(1,2)}(\lambda)$ converge a.s. as $k \to \infty$ to well-defined limits, that satisfy the following equations:

$$\begin{aligned} S^{(1,1)}(\lambda) &= g_\nu(\lambda)S^{(2)}(\lambda) - [1+\lambda g_\nu(\lambda)]\partial_\lambda S^{(1)}(\lambda) \\ &\quad + \alpha g_\nu(\lambda)[g_\nu(\lambda) + S^{(1)}(\lambda)]\int \frac{\rho_\Delta(\mathrm{d}t)t}{(1+tg_\nu(\lambda))^2}[t\,\partial_\lambda S^{(1)}(\lambda) - g_\nu(\lambda)], \\ S^{(1,2)}(\lambda) &= -[1+\lambda g_\nu(\lambda)][S^{(1,1)}(\lambda) + (1+\alpha)\partial_\lambda S^{(1)}(\lambda)] + g_\nu(\lambda)S^{(3)}(\lambda) \\ &\quad + \alpha g_\nu(\lambda)[(1+\alpha)g_\nu(\lambda) + S^{(1)}(\lambda) + S^{(2)}(\lambda)]\int \frac{\rho_\Delta(\mathrm{d}t)t}{(1+tg_\nu(\lambda))^2}[t\,\partial_\lambda S^{(1)}(\lambda) - g_\nu(\lambda)]. \end{aligned}$$

Lemma 5.6 (*Properties of $S^{(r)}$*)

Let $\alpha, \Delta > 0$. We focus mainly on $S^{(2)}(\lambda)$. We have:

(*i*) For every r, $S^{(r)}(\lambda)$ is a strictly increasing function of λ, and $\lim_{\lambda\to\infty} S^{(r)}(\lambda) = 0$.

(*ii*) For every $\lambda > \lambda_{\max}$, $S^{(2)}(\lambda) = -\alpha\Delta$ iff $\Delta \leq \Delta_c(\alpha)$ and $\lambda = 1$.

(*iii*) For every $\Delta > \Delta_c(\alpha)$, $\lim_{\lambda\downarrow\lambda_{\max}} S^{(2)}(\lambda) \in (-\alpha\Delta, 0)$.

Let us first see how item (*ii*) of Lemma 5.6 and eq. (5.33) end the proof of the eigenvalue transition. First, note that by the celebrated Weyl's interlacing inequalities [Wey49], we have:

$$\liminf_{p\to\infty} \lambda_1 \geq \lambda_{\max} \qquad \text{and} \qquad \limsup_{p\to\infty} \lambda_2 \leq \lambda_{\max}.$$

This implies that because the perturbation of the matrix is of rank one, *at most one* outlier eigenvalue will exist in the limit $p \to \infty$. By eq. (5.33), this outlier λ_1 exists if and only if it

[7] The almost sure convergence can be extended to all $r \in \mathbb{N}^\star$ but we will only use these values of r in the following.

satisfies, in the large $p \to \infty$ limit, the equation $S^{(2)}(\lambda_1) = -\alpha\Delta$. By item (ii) of Lemma 5.6, this is the case only for $\lambda_1 = 1$ and $\Delta \leq \Delta_c(\alpha)$, which ends the proof. □

A completely rigorous treatment of the previous arguments requires to state more precisely concentration results. Such a treatment has been made in [BGN11] in a close case (from which all the arguments transpose), and we refer to it for more details. Lemma 5.5, which is at the core of our proof, is proven in the following. We postpone the proof of Lemma 5.6 to Appendix D.3.1.

Proof of Lemma 5.5 – The essence of the computation originates from the derivation of Theorem 1.7 in [SB95], and is in essence a *cavity* computation, see Section 1.4. Note that $S_k^{(0)}(\lambda)$ converges a.s. to the Stieltjes transform $g_\nu(\lambda)$ as $k \to \infty$ by Theorem 1.7. For every $1 \leq i \leq p$, $\mathbf{w}_i$ denotes the i-th row of $\mathbf{W}$. We denote $\mathbf{y} = \boldsymbol{\xi}/\sqrt{\Delta p} - \mathrm{I}_p/\Delta$. Since $\mathbf{W}$ is independent of $\mathbf{y}$, we can denote $y_1, \cdots, y_p$ the eigenvalues of $\mathbf{y}$, and their empirical distribution converges a.s. to ρ_Δ as we know. We have in distribution:

$$\boldsymbol{\Gamma}_k^{(0)} = \frac{1}{k}\mathbf{W}^\intercal \mathbf{y}\, \mathbf{W} \overset{\mathrm{d}}{=} \frac{\alpha}{p}\sum_{i=1}^{p} y_i\, \mathbf{w}_i\, \mathbf{w}_i^\intercal.$$

For every i, we denote $\boldsymbol{\Gamma}_{k,i}^{(0)} \equiv (\alpha/p)\sum_{j(\neq i)}^{p} y_j\, \mathbf{w}_j\, \mathbf{w}_j^\intercal$. Note that $\boldsymbol{\Gamma}_{k,i}^{(0)}$ is independent of w_i. We start from the (trivial) decomposition, for every λ:

$$-\frac{1}{\lambda} = (\boldsymbol{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1} - \frac{1}{\lambda}\frac{\mathbf{W}^\intercal \mathbf{y}\, \mathbf{W}}{k}(\boldsymbol{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}. \tag{5.34}$$

We will make use of the Sherman-Morrison formula that gives the inverse of a matrix perturbed by a rank-one change:

$$(\mathbf{B} + t\boldsymbol{\omega}\boldsymbol{\omega}^\intercal)^{-1} = \mathbf{B}^{-1} - \frac{t}{1 + t\boldsymbol{\omega}^\intercal \mathbf{B}^{-1}\boldsymbol{\omega}}\mathbf{B}^{-1}\boldsymbol{\omega}\boldsymbol{\omega}^\intercal\mathbf{B}^{-1}. \tag{5.35}$$

Using it in eq. (5.34) yields:

$$-\frac{1}{\lambda} = -\frac{\alpha}{\lambda}\frac{1}{p}\sum_{i=1}^{p} y_i \frac{\mathbf{w}_i\mathbf{w}_i^\intercal(\boldsymbol{\Gamma}_{k,i}^{(0)} - \lambda \mathrm{I}_k)^{-1}}{1 + \frac{y_i}{k}\mathbf{w}_i^\intercal(\boldsymbol{\Gamma}_{k,i}^{(0)} - \lambda \mathrm{I}_k)^{-1}\mathbf{w}_i} + (\boldsymbol{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}. \tag{5.36}$$

Taking the trace of eq. (5.36), using the independence of $\mathbf{w}_i$ and $\boldsymbol{\Gamma}_{k,i}^{(0)}$, and the concentration $(1/k)\mathbf{w}_i^\intercal \mathbf{A}\mathbf{w}_i = (1/k)\mathrm{Tr}\mathbf{A}$ with high probability for large k, we obtain the following equation:

$$-\frac{1}{\lambda} = g_\nu(\lambda) - g_\nu(\lambda)\frac{\alpha}{\lambda}\int \rho_\Delta(\mathrm{d}t)\frac{t}{1 + t g_\nu(\lambda)}. \tag{5.37}$$

This is exactly Theorem 1.7! In the following, we will use very similar identities. A completely rigorous derivation of these would, however, require many technicalities to ensure in particular the concentration of all the involved quantities. It would exactly follow the proof of [SB95], and thus we do not repeat all the technicalities here. We can multiply eq. (5.36) by $\mathbf{W}^\intercal\mathbf{W}/k$, and take the trace:

$$-\frac{1}{\lambda}\frac{1}{k}\mathrm{Tr}\left[\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\right] = S_k^{(1)}(\lambda) - \frac{\alpha}{\lambda}\frac{1}{p}\sum_i y_i \frac{\frac{\mathbf{w}_i^\intercal}{\sqrt{k}}(\boldsymbol{\Gamma}_{k,i}^{(0)} - \lambda \mathrm{I}_k)^{-1}(\frac{1}{k}\sum_j \mathbf{w}_j\mathbf{w}_j^\intercal)\frac{\mathbf{w}_i}{\sqrt{k}}}{1 + \frac{y_i}{k}\mathbf{w}_i^\intercal(\boldsymbol{\Gamma}_{k,i}^{(0)} - \lambda \mathrm{I}_k)^{-1}\mathbf{w}_i}.$$

This implies that $S_k^{(1)}(\lambda)$ converges as $p, k \to \infty$ to a limit $S^{(1)}(\lambda)$, which satisfies:

$$-\frac{\alpha}{\lambda} = S^{(1)}(\lambda) - \frac{\alpha}{\lambda}\Big[\int \rho_\Delta(\mathrm{d}t)\frac{t}{1+tg_\nu(\lambda)}\Big](g_\nu(\lambda)+S^{(1)}(\lambda))$$

Using finally eq. (5.37), it is equivalent to:

$$S^{(1)}(\lambda) = g_\nu(\lambda)[\alpha - (1+\lambda g_\nu(\lambda))].$$

Multiplying eq. (5.36) by $(\mathbf{W}^\intercal\mathbf{W}/k)^2$ or $(\mathbf{W}^\intercal\mathbf{W}/k)^3$ yields, by the same analysis:

$$\begin{aligned}
S^{(2)}(\lambda) &= [\alpha(1+\alpha) - (1+2\alpha)(1+\lambda g_\nu(\lambda)) + (1+\lambda g_\nu(\lambda))^2]g_\nu(\lambda),\\
S^{(3)}(\lambda) &= [(\alpha+3\alpha^2+\alpha^3) - (1+5\alpha+3\alpha^2)(1+\lambda g_\nu(\lambda)) + (2+3\alpha)(1+\lambda g_\nu(\lambda))^2\\
&\qquad - (1+\lambda g_\nu(\lambda))^3]g_\nu(\lambda).
\end{aligned}$$

The convergence of $S_k^{(1,1)}(\lambda)$ and $S_k^{(1,2)}(\lambda)$ follows from the same analysis, as well as the equations they satisfy. We detail the derivation of the equation on $S^{(1,1)}(\lambda)$ and leave the derivation of the second equation for the reader. We multiply eq. (5.36) by $\mathbf{W}^\intercal\mathbf{W}/k$. To simplify the calculations, we make use of concentrations, and denote $\mathbf{F}_i \equiv (\mathbf{W}^\intercal\mathbf{W} - \mathbf{w}_i\mathbf{w}_i^\intercal)/k$, which is independent of $\mathbf{w}_i$. We obtain at leading order as $p \to \infty$:

$$\begin{aligned}
-\frac{\mathbf{W}^\intercal\mathbf{W}}{k\lambda} &= (\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\frac{\mathbf{W}^\intercal\mathbf{W}}{k} - \frac{\alpha}{\lambda}\frac{1}{p}\sum_{i=1}^p \frac{y_i g_\nu(\lambda)\mathbf{w}_i\mathbf{w}_i^\intercal}{1+y_i g_\nu(\lambda)}\\
&\quad - \frac{\alpha}{\lambda}\frac{1}{p}\sum_{i=1}^p \frac{y_i}{1+y_i g_\nu(\lambda)}\mathbf{w}_i\mathbf{w}_i^\intercal(\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}\mathbf{F}_i.
\end{aligned}$$

We multiply this equation by $(\mathbf{\Gamma}_k^{(0)} - \lambda\mathrm{I}_k)^{-1}$ and use eq. (5.35) in the form:

$$(\mathbf{\Gamma}_k^{(0)} - \lambda\mathrm{I}_k)^{-1} = (\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1} - (\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}\frac{y_i\mathbf{w}_i\mathbf{w}_i^\intercal}{1+y_i g_\nu(\lambda)}(\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}.$$

Using again simple concentration, this yields at leading order:

$$\begin{aligned}
&-\frac{\mathbf{W}^\intercal\mathbf{W}}{k\lambda}(\mathbf{\Gamma}_k^{(0)} - \lambda\mathrm{I}_k)^{-1} =\\
&(\mathbf{\Gamma}_k^{(0)} - \lambda\mathrm{I}_k)^{-1}\frac{\mathbf{W}^\intercal\mathbf{W}}{k}(\mathbf{\Gamma}_k^{(0)} - \lambda\mathrm{I}_k)^{-1} - \frac{\alpha}{\lambda}\frac{1}{p}\sum_{i=1}^p \frac{y_i\mathbf{w}_i\mathbf{w}_i^\intercal(\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}\mathbf{F}_i(\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}}{1+y_i g_\nu(\lambda)}\\
&+ \frac{\partial_\lambda S^{(1)}(\lambda)}{\lambda}\frac{\alpha}{p}\sum_{i=1}^p \frac{y_i^2\mathbf{w}_i\mathbf{w}_i^\intercal(\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}}{(1+y_i g_\nu(\lambda))^2} - \frac{\alpha}{\lambda}\frac{1}{p}\sum_{i=1}^p \frac{y_i g_\nu(\lambda)\mathbf{w}_i\mathbf{w}_i^\intercal(\mathbf{\Gamma}_{k,i}^{(0)} - \lambda\mathrm{I}_k)^{-1}}{(1+y_i g_\nu(\lambda))^2}.
\end{aligned}$$

We finally multiply this equation by $\mathbf{W}^\intercal\mathbf{W}/k$ and take the trace. Using again usual concentrations we reach:

$$\begin{aligned}
-\frac{S^{(2)}(\lambda)}{\lambda} &= S^{(11)}(\lambda) - \frac{\alpha}{\lambda p}\sum_{i=1}^p \frac{y_i[S^{(11)}(\lambda) + \partial_\lambda S^{(1)}(\lambda)]}{1+y_i g_\nu(\lambda)}\\
&\quad + \frac{\partial_\lambda S^{(1)}(\lambda)}{\lambda}\frac{\alpha}{p}\sum_{i=1}^p \frac{y_i^2[g_\nu(\lambda)+S^{(1)}(\lambda)]}{(1+y_i g_\nu(\lambda))^2} - \frac{\alpha}{\lambda}\frac{1}{p}\sum_{i=1}^p \frac{y_i g_\nu(\lambda)}{(1+y_i g_\nu(\lambda))^2}[g_\nu(\lambda)+S^{(1)}(\lambda)].
\end{aligned} \tag{5.38}$$

We now take the limit $p \to \infty$ and use Theorem 1.7 in the form:

$$\frac{\alpha}{\lambda}\int \rho_\Delta(\mathrm{d}t)\frac{t}{1+tg_\nu(\lambda)} = 1 + \frac{1}{\lambda g_\nu(\lambda)}.$$

Inserting this into eq. (5.38) along with some trivial algebra yields the sought identity:

$$\begin{aligned} S^{(1,1)}(\lambda) = {} & g_\nu(\lambda)S^{(2)}(\lambda) - [1+\lambda g_\nu(\lambda)]\partial_\lambda S^{(1)}(\lambda) \\ & + \alpha g_\nu(\lambda)[g_\nu(\lambda)+S^{(1)}(\lambda)]\int \frac{\rho_\Delta(\mathrm{d}t)t}{(1+tg_\nu(\lambda))^2}[t\,\partial_\lambda S^{(1)}(\lambda) - g_\nu(\lambda)], \end{aligned}$$

which is what we aimed to show. Performing the same analysis for $S^{(1,2)}(\lambda)$ ends the proof. □

The proof of correlation of the leading eigenvector is derived on the same principles, and is given in Appendix D.3.2. All together, this ends the proof of Theorem 5.3.

On the nature of the transition – As was already noticed in previous works (see e.g. a related remark in [BGN11]), the existence of a transition in the largest eigenvalue and the corresponding eigenvector for a large matrix of the type $M + \theta P$ (with P of finite rank and $\theta > 0$) depends on the decay of the asymptotic spectral density of M at the right edge of its bulk. For a power-law decay, there can be either no transition, a transition in the largest eigenvalue and the corresponding eigenvector, or a transition in the largest eigenvalue but not in the corresponding eigenvector. The situation in our setting is somewhat more involved, as both the bulk and the spike depend on the parameter Δ, and they are not independent (they are correlated via the matrix $\mathbf{W}$). However, this intuition remains true: if we do not show and use it explicitly, the decay of the density of $\mu(\alpha, \Delta)$ at the right edge is of the type $(\lambda_{\max} - \lambda)^{1/2}$, which is the hidden feature that is responsible for a transition both in the largest eigenvalue and the corresponding eigenvector that what we show in Theorem 5.3.

Perspectives on Chapter 5

This chapter presented a detailed analysis of the influence of data structure on optimal learning in the spiked-matrix model. We modeled the data structure by *generative priors*, and were motivated by comparing our findings to the case of *sparse* data, for which the underlying structure induces computational gaps. We detailed the two main conclusions of our analysis in Section 5.1.4: first, in contrast to the sparse case, there is no algorithmic gap with random generative priors. And secondly, we have designed a spectral algorithm (LAMP, Algorithm 5) that reaches the optimal weak-recovery threshold and that we can analyze both using the asymptotic analysis of AMP and random matrix theory.

While this chapter is based mainly on [ALM+20], part of the theoretical machine learning community has very recently gained interest in understanding the role of data structure in learning, especially in neural networks. As highlighted in [Zde20], data structure is one of the three key elements needed to build a theory of neural network learning, along with the network architecture and the optimization algorithms. It is therefore quite natural to see more and more recent works in the same general line as this chapter. To name a few (from collaborators), the reader can refer to [ALB+20] for the phase retrieval problem, or [GMKZ20] which introduced a similar model called *hidden manifold*, that allowed to study the influence of structure in a very flexible manner. While the majority of these works rely on synthetic (i.e. random) datasets, recent theoretical and empirical studies are indicating that a large part of their results remains valid for trained generative priors [GRM+20]. All these contributed to create an exciting and rapidly growing line of work, and one aim of this chapter was to show that many tools of

statistical physics applied to inference are particularly suited to build the premises of a theory of the role of data structure in learning.

Chapter 6

Phase retrieval: theoretical transitions and efficient algorithms

"Quand les physiciens nous demandent la solution d'un problème, ce n'est pas une corvée qu'ils nous imposent, c'est nous au contraire qui leur devons des remerciements."

Henri Poincaré, La Valeur de la Science – Chapitre V: L'Analyse et la Physique (1911).

Disclaimer – We end our tour of high-dimensional inference problems by one of the most renowned models in this class: *phase retrieval*. This dense chapter merges items from a large part of our toolbox: the replica method and the means to prove its results, message-passing algorithms to study hardness, but also the TAP picture, and the results of Chapter 2, to understand spectral algorithms to solve phase retrieval. It is mainly based on two published works [MLKZ20, MKLZ21], which focused respectively on the fundamental limits of phase retrieval, and on the design of optimal spectral methods to solve it.

Recall our mathematical notation: we let $\beta \in \{1, 2\}$ for respectively real ($\mathbb{K} = \mathbb{R}$) and complex ($\mathbb{K} = \mathbb{C}$) variables.

6.1 The phase retrieval problem

Consider the reconstruction problem of a signal $\mathbf{X}^* \in \mathbb{K}^n$ from m observations of its modulus

$$Y_\mu = \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_{\mu i} X_i^* \Big|, \qquad \mu = 1, \cdots, m, \tag{6.1}$$

where the $m \times n$ *sensing* matrix $\mathbf{\Phi} \in \mathbb{K}^{m \times n}$ is known. More generally, measurements can be a noisy function of the modulus, for example by an additive Gaussian noise, or by a Poisson noise channel. This inverse problem, known in the literature under the umbrella of *phase retrieval* arises in a large set of problems ranging from signal processing [Fie82, UE88, DLM+15] to statistical estimation [CLS15b, JEH15], optics, X-ray crystallography, astronomy or microscopy [SEC+15], where detectors can often only measure information about the amplitude of signals, and lose all information about its phase, cf. Fig. 6.1. Phase retrieval is also a textbook example of a learning problem with a highly non-convex loss landscape [NJS15, SQW18, HLV18]. It moreover falls into the large category of generalized linear models (GLMs), that we introduced in Section 1.1.

Phase retrieval and generalized linear models – Let us first precise the class of models studied in this chapter. We assume that the (random) sensing matrix $\mathbf{\Phi}$ satisfies *right-rotation invariance*, in the sense of Model R. The observations are then generated by a Bayes-optimal GLM similar to the ones we encountered in the previous chapters. Precisely, the signal $\mathbf{X}^*$

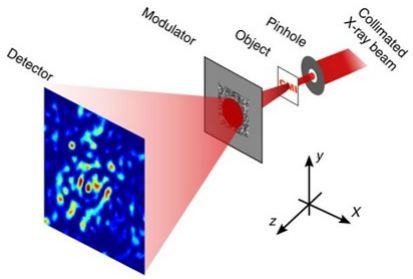


FIGURE 6.1: Schematic of a Coherent Modulation Imaging set-up, a typical optical experiment in which one needs to retrieve measurements from the observation of the modulus of their projections (here by a modulator), which can be modeled by eq. (6.2). Picture from [ZCM+16].

to recover is drawn using a factorized prior distribution P_0 with zero mean and variance $\rho \equiv \mathbb{E}_{P_0}[|x|^2] > 0$. The observations are generated as:

$$Y_\mu \sim P_{\text{out}}\Big(\cdot \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_{\mu i} X_i^\star \Big), \quad 1 \le \mu \le m, \tag{6.2}$$

and we assume that the prior P_0 and the "channel" distribution $P_{\text{out}}(y|z)$ are known. Eq. (6.2) defines the very general class of *Generalized Linear Models* (GLMs), cf. Section 1.1. The present chapters focuses on *phase retrieval* problems:

Definition 6.1 (*Phase retrieval*)

We generically denote as *phase retrieval* GLMs of the type of eq. (6.2) in which we assume that $P_{\text{out}}(y|z)$ is a function of $|z|$, and in which the prior distribution P_0 is also symmetric: $P_0(x) = P_0(|x|)$.

For instance, for Gaussian additive noise one has $P_{\text{out}}(y|z) = \mathcal{N}_1(y; |z|^2, \Delta)$, while the noiseless case corresponds to the limit $\Delta \downarrow 0$: $P_{\text{out}}(y|z) = \delta(y - |z|^2)$. Finally, we consider a now-standard *high-dimensional (thermodynamic) limit*, in which $n, m \to \infty$ with $m/n \to \alpha > 0$. All in all, the model we consider is very generic, and encompasses e.g. phase retrieval with unitary matrices (and in some extent *Fourier phase retrieval* as we will see), or Gaussian phase retrieval.

Related literature – A great amount of work is present both in the statistical physics and in the information theory literature for different assumptions on the matrix $\mathbf{\Phi}$. The asymptotic optimal performances (both information-theoretic and algorithmic) of generic GLMs with right-rotation invariance sensing matrices were conjectured using the non-rigorous replica method of statistical physics in [Kab08a, TK20], and these results were proven for Gaussian matrices in [BKM+19]. This analysis was later non-rigorously extended to the case of non-separable prior distributions [ALB+20]. Specifically for the phase retrieval problem, the limits of weak-recovery were analyzed for Gaussian matrices $\mathbf{\Phi}$ in [MM19, LAL19, LL20], and for column-unitary $\mathbf{\Phi}$ in [DMM20, DBMM20, MDX+21][1]. As we know, the Bayes-optimal estimation can be summarized in the study of the *posterior probability* of $\mathbf{x}$ given the observations $\mathbf{Y}$ and the matrix $\mathbf{\Phi}$:

$$\mathbb{P}(\mathbf{x}|\mathbf{Y}, \mathbf{\Phi}) \equiv \frac{1}{\mathcal{Z}_n(\mathbf{Y})} \prod_{i=1}^{n} P_0(x_i) \prod_{\mu=1}^{m} P_{\text{out}}\Big(Y_\mu \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_{\mu i} x_i\Big). \tag{6.3}$$

We will generically denote by $\langle \cdot \rangle$ the average with respect to the posterior probability (6.3).

[1] All these works (and this chapter) consider *random* instances of phase retrieval, while there is another important literature that studies *injectivity* of phase retrieval. For instance, the perfect recovery of an arbitrary (rather than typical) signal in complex phase retrieval has been studied in [BCMN14], and the thresholds are very different.

On the algorithmic side, the (generalized) *vector approximate message-passing* (G-VAMP) algorithm [SRF16, RSF17] (cf. Section 1.4 and Chapter 2) has been conjectured to achieve the optimal polynomial-time recovery for rotationally (unitarily) invariant matrices[2]. Despite its properties, G-VAMP is a numerically very cumbersome algorithms. It is therefore of great interest to investigate more robust and computationally cheaper algorithms that could achieve similar performances. A natural class of such algorithms are *spectral methods*, i.e. methods based on the leading eigenvector of a matrix constructed from the observations. Their output can be used as informative initializations for local gradient-based optimization algorithms, and can induce a jump in the accuracy achieved at a reasonable computational cost. Such techniques have already been applied e.g. in optical systems [VDG21]. In the context of phase retrieval, the performance of these spectral methods has been rigorously analyzed for Gaussian [MM19, LAL19, MTV20, LL20] and unitary [DBMM20, DB20, MDX+21] sensing matrices.

Organization of the chapter and main contributions – The present chapter extends all these past results by considering arbitrary matrices with orthogonal or unitary invariance properties, encapsulating all the cases described above into a single framework. Our main contributions are twofold:

(I) First, we derive sharp asymptotics for the lowest possible estimation error achievable statistically and algorithmically, locate the phase transitions for weak- and full-recovery as a function of the singular values of the matrix $\mathbf{\Phi}$ and also discuss the existence of a statistical-to-algorithmic gap. Our main contributions on this matter are:

- In Section 6.2, we use the replica method to derive a unified single-letter formula for the performance of the Bayes-optimal estimator for generic real or complex GLMs defined by eq. (6.2) in the thermodynamic limit, and for right-rotationally invariant $\mathbf{\Phi}$. We then prove this conjecture in two particular cases. First, when the distribution P_0 is Gaussian and $\mathbf{\Phi} = \mathbf{WB}$ is the product of a Gaussian matrix $\mathbf{W}$ with an "arbitrary" matrix $\mathbf{B}$. Second, for a Gaussian matrix $\mathbf{\Phi}$ (real or complex) with any separable distribution P_0. These are non-trivial extensions of the the proofs of [BKM+19, BM19a] and [BMMK18]. We also argue that message-passing algorithms, here G-VAMP, achieves the optimal performance reachable in polynomial time among a large class of algorithms, and describe its iterations.
- In Section 6.3, we analytically characterize (as a function of the singular values distribution of $\mathbf{\Phi}$) the *algorithmic weak-recovery* threshold $\alpha_{\mathrm{WR,Algo}}$ above which better-than-random inference reconstruction of $\mathbf{X}^*$ is possible in polynomial time. We also derive an explicit formula for the *information-theoretic full recovery* threshold $\alpha_{\mathrm{FR,IT}}$ above which full reconstruction of $\mathbf{X}^*$ (i.e. perfect recovery up to the possible rank deficiency of $\mathbf{\Phi}$) is statistically possible, as a function of the singular values distribution of $\mathbf{\Phi}$.

Our findings for the statistical and algorithmic thresholds are summarized in Table 6.1, for different real and complex ensembles of $\mathbf{\Phi}$. Entries in bold emphasize new results obtained in this chapter, filling a gap between the different previous works in the phase retrieval literature.

(II) Secondly, in Section 6.4 we design (conjecturally) *optimal* spectral methods for the phase retrieval problem in the aforementioned limit, for the very generic class of right-rotationally invariant sensing matrices. Most importantly, in contrast to previous works our approach is completely *automated*, in the sense that our spectral methods are conjectured to be optimal, and their derivation does not require optimization over any hyperparameter. The constructiveness of our approach gives more weight to our optimality conjecture, as we do not restrict

[2] To test the performance of the G-VAMP algorithm, we used the TrAMP library [BAKZ20] that provides an open-source implementation.

Matrix ensemble and value of β	$\alpha_{\text{WR,Algo}}$	$\alpha_{\text{FR,IT}}$	$\alpha_{\text{FR,Algo}}$
Real Gaussian $\mathbf{\Phi}$ ($\beta = 1$)	0.5 [MM19, LAL19]	1 [CT06]	$\simeq 1.12$ [BKM+19]
Complex Gaussian $\mathbf{\Phi}$ ($\beta = 2$)	1 [MM19, LAL19]	**2**	$\simeq$ **2.027**
Real column-orthogonal $\mathbf{\Phi}$ ($\beta = 1$)	**1.5**	1 [CT06]	$\simeq$ **1.584**
Complex column-unitary $\mathbf{\Phi}$ ($\beta = 2$)	2 [MP17, MDX+21]	**2**	$\simeq$ **2.265**
$\mathbf{\Phi} = \mathbf{W}_1\mathbf{W}_2$ ($\beta = 1$, aspect ratio γ)	$\gamma/(2(1+\gamma))$ [ALB+20]	$\min(1,\gamma)$ [CT06]	Thm. 6.2 [ALB+20]
$\mathbf{\Phi} = \mathbf{W}_1\mathbf{W}_2$ ($\beta = 2$, aspect ratio γ)	$\gamma/(\mathbf{1}+\gamma)$	$\min(\mathbf{2}, \mathbf{2}\gamma)$	**Thm. 6.2**
$\mathbf{\Phi}$, $\beta \in \{1,2\}$, $\text{rk}[\mathbf{\Phi}^\dagger\mathbf{\Phi}]/n = r$	**Eq. (6.13)**	$\beta\mathbf{r}$	**Conj. 6.1**
Gauss. $\mathbf{\Phi}$, $\beta \in \{1,2\}$, symm. P_0, P_{out}	Eq. (6.12) [MM19, LAL19]	**Thm. 6.2**	**Thm. 6.2**
$\mathbf{\Phi}$, $\beta \in \{1,2\}$, symm. P_0, P_{out}	**Eq. (6.11)**	**Conj. 6.1**	**Conj. 6.1**

TABLE 6.1: Values of the algorithmic weak recovery, information-theoretic full recovery, and algorithmic full recovery thresholds for several random matrix ensembles. When the ensemble of $\mathbf{\Phi}$ is not specified, we consider any right-orthogonally (unitarily) invariant ensemble in the sens of Model R. The last two lines are given for any symmetric (cf Def. 6.1) prior P_0 and channel P_{out}, while all other results are for Gaussian P_0 and a noiseless phase retrieval channel. We reference results of this chapter when the value is not given by a closed-form expression, but can be computed from the formulas herein. In some particular ensembles, we have numerically analyzed these equations in Section 6.5. The new results of this chapter are written in bold blue, and we give references to papers in which the previously known thresholds were computed.

to a specific family of spectral methods. In designing these methods, we unify three different approaches: a "pedestrian" optimization of the preprocessing function (the approach of the aforementioned previous works), the linearization of message-passing algorithms, and a *Bethe Hessian* analysis.

Finally, in Section 6.5, we numerically investigate our two contributions. In particular, we uncover interesting transition phenomena in the eigenvalues of the spectral methods, and we numerically establish the existence or absence of a statistical-to-algorithmic gap for many ensembles of $\mathbf{\Phi}$ in noiseless phase retrieval, for which such an analysis was lacking.

Important notation – Recall that for $m \geq n$, a matrix $\mathbf{A} \in \mathbb{K}^{m\times n}$ is said to be *column-orthogonal (unitary)* if $\mathbf{A}^\dagger\mathbf{A} = \mathrm{I}_n$. For $x, y \in \mathbb{K}$, we define a 'dot product' as $x \cdot y \equiv xy$ if $\mathbb{K} = \mathbb{R}$ and $x \cdot y \equiv \text{Re}[\bar{x}y]$ if $\mathbb{K} = \mathbb{C}$. In particular $x \cdot x = |x|^2$. The Gaussian measure $\mathcal{N}_\beta(0,1)$ is defined as $\mathcal{D}_\beta z \equiv (2\pi/\beta)^{-\beta/2} \exp(-\beta|z|^2/2)\,\mathrm{d}z$. ν will denote the asymptotic spectral density of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ and we designate $\langle f(\lambda)\rangle_\nu \equiv \int \nu(\mathrm{d}\lambda) f(\lambda)$ the linear statistics of ν.

6.2 Optimal estimation in GLMs with structured data

6.2.1 Replica free entropy and how to prove it

Recall that we placed ourselves in a setting known as *Bayes-optimal*: the statistically optimal estimator $\hat{\mathbf{X}}$ minimizing the mean-squared error $\text{mse}(\hat{\mathbf{X}}) \equiv ||\hat{\mathbf{X}} - \mathbf{X}^\star||_2^2$ is therefore simply given by the posterior mean $\hat{\mathbf{X}}_{\text{opt}} = \langle \mathbf{x} \rangle$, where the posterior distribution is given in eq. (6.3). As we know, while exact sampling from the posterior is intractable for large values of $n, m \in \mathbb{N}^*$, the replica method allows us to access important asymptotic quantities, such as the *free entropy*[3]. It fully characterizes the asymptotic performance of the Bayes-optimal estimator $\hat{\mathbf{X}}_{\text{opt}}$ in high dimensions via the I-MMSE theorem [GSV05].

[3]It is directly related to the mutual information between $\mathbf{X}^*$ and $\mathbf{Y}$, see Chapter 5 in which we adopted the mutual-information convention

The first result of this chapter is a single-letter formula for the asymptotic free entropy for right-rotationally invariant sensing (or data) matrices $\mathbf{\Phi}$[4], in the limit $n, m \to \infty$ with $m/n \to \alpha > 0$.

Conjecture 6.1 (*Replica-symmetric conjecture for generic GLMs*)

Under the assumptions above, the asymptotic free entropy for the posterior distribution defined in eq. (6.3) with right-orthogonally (unitarily) invariant sensing matrix $\mathbf{\Phi}$ is:

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{\mathbf{Y},\mathbf{\Phi}} \ln \mathcal{Z}_n(\mathbf{Y}) = \sup_{q_x\in[0,\rho]} \sup_{q_z\in[0,Q_z]} [I_0(q_x) + \alpha I_{\text{out}}(q_z) + I_{\text{int}}(q_x, q_z)], \tag{6.4}$$

where

$$\begin{cases} I_0(q_x) & \equiv \inf_{\hat{q}_x\geq 0} \Big[-\dfrac{\beta\hat{q}_x q_x}{2} + \mathbb{E}_\xi \mathcal{Z}_0(\sqrt{\hat{q}_x}\xi, \hat{q}_x) \ln \mathcal{Z}_0(\sqrt{\hat{q}_x}\xi, \hat{q}_x)\Big], \\ I_{\text{out}}(q_z) & \equiv \inf_{\hat{q}_z\geq 0} \Bigg[-\dfrac{\beta\hat{q}_z q_z}{2} - \dfrac{\beta}{2}\ln(\hat{Q}_z + \hat{q}_z) + \dfrac{\beta\hat{q}_z}{2\hat{Q}_z} \\ & \quad + \mathbb{E}_\xi \displaystyle\int_{\mathbb{R}} \mathrm{d}y\ \mathcal{Z}_{\text{out}}\Big(y; \sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi, \frac{1}{\hat{Q}_z+\hat{q}_z}\Big) \ln \mathcal{Z}_{\text{out}}\Big(y; \sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi, \frac{1}{\hat{Q}_z+\hat{q}_z}\Big)\Bigg], \\ I_{\text{int}}(q_x, q_z) & \equiv \inf_{\gamma_x,\gamma_z\geq 0} \Big[\dfrac{\beta}{2}(\rho - q_x)\gamma_x + \dfrac{\alpha\beta}{2}(Q_z - q_z)\gamma_z - \dfrac{\beta}{2}\langle \ln(\rho^{-1} + \gamma_x + \lambda\gamma_z)\rangle_\nu\Big] \\ & \quad -\dfrac{\beta}{2}\ln(\rho - q_x) - \dfrac{\beta q_x}{2\rho} - \dfrac{\alpha\beta}{2}\ln(Q_z - q_z) - \dfrac{\alpha\beta q_z}{2Q_z}. \end{cases}$$

We defined $Q_z \equiv \rho\langle\lambda\rangle_\nu/\alpha$ and $\hat{Q}_z \equiv 1/Q_z$, $\xi \sim \mathcal{N}_\beta(0,1)$ and the following auxiliary functions:

$$\mathcal{Z}_0(b, a) \equiv \int_{\mathbb{K}} \mathrm{d}z\, P_0(z) e^{-\frac{\beta}{2}a|z|^2 + \beta b\cdot z}, \qquad \mathcal{Z}_{\text{out}}(y; \omega, v) \equiv \mathbb{E}_z[P_{\text{out}}(y|\sqrt{v}z + \omega)], \tag{6.5}$$

with $z \sim \mathcal{N}_\beta(0,1)$. Moreover, the asymptotic MMSE achieved by the Bayes-optimal estimator is equal to $\rho - q_x^*$, with q_x^* the solution of the above extremization problem:

$$\lim_{n\to\infty} \text{MMSE} = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\|\mathbf{X}^* - \hat{\mathbf{X}}_{\text{opt}}\|^2 = \rho - q_x^*. \tag{6.6}$$

This formula is derived in Appendix B.2 using the heuristic (hence the *conjecture*) replica method that we introduced in Section 1.3.1. Given its generality, the calculation of Appendix B.2 also serves in this thesis as a textbook example of application of the replica method, along with the one for the committee machine presented in Appendix B.1. It extends the formula from [TK20] to complex signals $\mathbf{X}^*$ and sensing matrices $\mathbf{\Phi}$. In particular, it also holds in the case of complex matrices $\mathbf{\Phi}$ and real signal $\mathbf{X}^\star$, by adding a constraint on the imaginary part of $\mathbf{X}^\star$ in P_0. It also encompasses the case of sparse signals, which is of wide interest in the compressive sensing literature [Don06, DMM09, KMS+12, SR14, KMTZ14].

Proving Conjecture 6.1 is a challenging open problem. We provide a significant step by doing so for a broad class of likelihoods P_{out} and in two settings: a restricted signal distribution P_0 and a broad class of real and complex likelihoods and sensing matrices $\mathbf{\Phi}$, or a broad class of prior distribution P_0 and (real or complex) Gaussian $\mathbf{\Phi}$. To state the theorem, we rewrite the model

[4] With respect to Model R we assume the following, which is (slightly) stronger: the large deviations of the spectral density of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ should happen in the scale $n^{1+\epsilon}$ for an $\epsilon > 0$. This condition was not precised in the replica calculation of [TK20] for real matrices. In practice, in classical orthogonally (unitarily)-invariant random matrix ensembles, we often have $\epsilon = 1$.

of eq. (6.2) as:

$$Y_\mu = \varphi_{\text{out}}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{\mu i}X_i^\star, A_\mu\Big), \quad 1 \le \mu \le m, \tag{6.7}$$

where $(A_\mu)_{\mu=1}^m \in \mathbb{K}^m$ are i.i.d. random variables with (known) distribution P_A accounting for a possible noise, φ_{out} is the observation channel and $\mathbf{\Phi}$ is a random matrix with elements in $\mathbb{K}$. $P_{\text{out}}(\cdot|z)$ is then the PDF associated to the stochastic function $\varphi_{\text{out}}(z, A)$.

Theorem 6.2 (*Replica-symmetric formula*)

Let us define a set of hypotheses:

$(H0)$ $\varphi_{\text{out}} : \mathbb{K}^2 \to \mathbb{R}$ is $\mathcal{C}^2$, and $(z, a) \mapsto (\varphi_{\text{out}}(z, a), \partial_z\varphi_{\text{out}}(z, a), \partial_z^2\varphi_{\text{out}}(z, a))$ is bounded.

(h1) P_0 is a centered Gaussian distribution, without loss of generality $P_0 = \mathcal{N}_\beta(0, 1)$.

(h2) $\mathbf{\Phi}$ is distributed as $\mathbf{\Phi} \overset{\text{d}}{=} \mathbf{W}\mathbf{B}/\sqrt{p}$, with $\mathbf{W} \in \mathbb{K}^{m\times p}$ an i.i.d. standard Gaussian matrix, and $\mathbf{B} \in \mathbb{K}^{p\times n}$ an arbitrary matrix (random or deterministic), independent of $\mathbf{W}$. Moreover, as $n \to \infty$, $p/n \to \delta > 0$.

(h3) The ESD of $\mathbf{B}^\dagger\mathbf{B}/n$ weakly converges (a.s.) to a compactly-supported measure $\nu_B \neq \delta_0$. Moreover, there exists $\lambda_{\max} > 0$ such that a.s. $\lambda_{\max}(\mathbf{B}^\dagger\mathbf{B}/n) \to_{n\to\infty} \lambda_{\max}$.

$(h'1)$ P_0 has a finite second moment, and $\Phi_{\mu i} \overset{\text{i.i.d}}{\sim} \mathcal{N}_\beta(0, 1)$.

Assume that all $(H0)$,(h1),(h2),(h3) or that all $(H0)$,$(h'1)$ stand. Then Conjecture 6.1 holds with ν the LSD of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$[5].

The proof is based on the adaptive interpolation method[6]. Its main strategy is sketched in the following, and shares similarities with the committee machine that we studied in Chapter 4. In particular, Theorem 6.2 allows to rigorously compute the asymptotic minimum mean-squared error (MMSE) achieved by the Bayes-optimal estimator.

Remark 6.1 (*Relaxing the hypotheses*)

Following arguments of [BKM+19] and Chapter 4, $(H0)$ can be relaxed to continuity a.e. and the existence of moments of φ_{out}, so that our theorem also covers noiseless phase retrieval.

As with the previous high-dimensional models we studied in Part II, the replica formula reduces the high-dimensional computation of the MMSE to a simple low-dimensional extremization problem. The MMSE as a function of the sample complexity α can be readily computed from eqs. (6.4) and (6.6) for a given signal distribution P_0 (determining I_0), likelihood P_{out} (determining I_{out}) and spectral density ν (determining I_{int}).

Proof strategy for Theorem 6.2

Let us work under $(H0)$,(h1),(h2),(h3). We will simply sketch the main steps of the proof strategy for the free energy. More details (e.g. going from the free entropy to the MMSE characterization), as well as the extension to hypotheses $(H0)$,$(h'1)$, can be found in [MLKZ20].

First, we simplify the conjectured expression of the free entropy of Conjecture 6.1 using the particular form of the prior P_0 and of the sensing matrix $\mathbf{\Phi}$. Finally, using (h1),(h2),(h3) and a

[5] The rigorous statement on the limit of the MMSE requires adding a side information channel with arbitrarily small signal, as is detailed in [MLKZ20].

[6] In Theorem 6.2, we rely on some Gaussianity, either in the prior or in the data matrix. This is a not specific to our setting, but rather a fundamental limitation of the adaptive interpolation method used for the proof.

proof similar to the one of [BKM+19, AMB+19], we give a rigorous derivation of this simplified expression. Note that with respect to the analysis of [BKM+19, AMB+19], there are two main novelties in our setting:

(*i*) The sensing matrix $\mathbf{\Phi}$ is not i.i.d. but has a well-controlled structure, see (h2) and (h3).

(*ii*) The variables can be complex numbers. We will argue that the arguments generalize to this case. The physical reason of this generalization is that even in the complex setting, the overlap will concentrate on a real positive number, as a consequence of Bayes-optimality.

First, we note that we can simplify the replica conjecture under the considered hypotheses:

Proposition 6.3 (*Simplified replica-symmetric formula*)

Under (*H*0),(h1),(h2),(h3), the replica conjecture 6.1 for the free entropy $f_n \equiv n^{-1}\mathbb{E}\ln \mathcal{Z}_n(\mathbf{Y})$ is equivalent to:

$$\lim_{n\to\infty} f_n = \sup_{\hat{q}\geq 0}\inf_{q\in[0,Q_z]}\Big[\frac{\beta\hat{q}}{2}(\mathbb{E}_{\nu_B}[X]-\delta q) - \frac{\beta}{2}\mathbb{E}_{\nu_B}\ln(1+\hat{q}X) + \alpha\Psi_{\text{out}}(q)\Big], \tag{6.8}$$

with $Q_z = \mathbb{E}_{\nu_B}[X]/\delta$ and Ψ_{out} defined in terms of the auxiliary functions of eq. (6.5):

$$\Psi_{\text{out}}(q) \equiv \int_{\mathbb{K}} \mathcal{D}_\beta \xi \int_{\mathbb{R}} \mathrm{d}y\ \mathcal{Z}_{\text{out}}(y;\sqrt{q}\xi, Q_z - q)\ln \mathcal{Z}_{\text{out}}(y;\sqrt{q}\xi, Q_z - q).$$

The proof of Proposition 6.3 is quite technical and is detailed in [MLKZ20]. Proposition 6.3 presents the replica-symmetric formula in a manner closer to its usual statement for Gaussian matrices (e.g. Theorem 4.1 for the committee machine), as a single supinf involving an order parameter q playing the role of the *overlap*. To finish the proof of the free entropy statement of Theorem 6.2, we therefore just need to show:

Lemma 6.4

Under the assumptions of Proposition 6.3, the limit of the free entropy $f_n \equiv n^{-1}\mathbb{E}\ln \mathcal{Z}_n(\mathbf{Y})$ is given by eq. (6.8).

Let us now briefly describe our strategy to prove Lemma 6.4. The main idea is to reduce the problem to a Generalized Linear Model with a Gaussian sensing matrix, but a non-i.i.d. prior. We make use of the "SVD" decomposition of $\mathbf{B}/\sqrt{n} = \mathbf{USV}^\dagger$, with $\mathbf{U} \in \mathcal{U}_\beta(p)$, $\mathbf{V} \in \mathcal{U}_\beta(n)$, and $\mathbf{S} \in \mathbb{R}^{p\times n}$ a pseudo-diagonal matrix with positive elements. Leveraging on the gaussianity of P_0 (hypothesis (h1)), and that $\mathbf{W}$ is an i.i.d. Gaussian matrix independent of $\mathbf{B}$, one can see that our estimation problem is formally equivalent to an usual Generalized Linear Model with m measurements, a signal of dimension p, and a Gaussian i.i.d. sensing matrix. A key feature is that here the prior distribution on the data $\mathbf{Z}^* \in \mathbb{K}^p$ is defined as

- If $\delta \leq 1$, for every $k \in \{1,\cdots,p\}$, Z_k^* is distributed as $S_k X_k^*$ with $X_k^* \overset{\text{i.i.d.}}{\sim} P_0$.
- If $\delta \geq 1$, for every $k \in \{1,\cdots,n\}$, Z_k^* is distributed as $S_k X_k^*$ with $X_k^* \overset{\text{i.i.d.}}{\sim} P_0$, while for every $k \in \{n+1,\cdots,p\}$, Z_k^* is almost surely 0.

More precisely, we can define rigorously the prior $P_0^{(\mathbf{S})}$ described above by its linear statistics. For any continuous bounded function $g : \mathbb{K}^p \to \mathbb{R}$, one has:

$$\int_{\mathbb{K}^p} P_0^{(\mathbf{S})}(\mathrm{d}\mathbf{z})g(\mathbf{z}) \equiv \int_{\mathbb{K}^n}\Big\{\prod_{i=1}^n P_0(\mathrm{d}x_i)\Big\}g(\{\mathrm{I}[k\leq n]S_k x_k\}_{k=1}^p). \tag{6.9}$$

Algorithm 6: Generalized Vector Approximate Message Passing

Data: Sensing matrix $\mathbf{\Phi}/\sqrt{n} = \mathbf{USV}^\dagger$, outputs $\{y_\mu\}_{\mu=1}^m$, a number of iterations T.
Result: An estimate $\hat{\mathbf{x}}$ of $\mathbf{X}^*$.
Randomly initialize all variables;
for $t = 1, \cdots, T$ **do**

(Denoising $\boldsymbol{x}$)	*(Denoising $\boldsymbol{z} \equiv \frac{1}{\sqrt{n}}\mathbf{\Phi}\boldsymbol{x}$)*
$\hat{\mathbf{x}}_1^t = g_{x1}(\mathbf{T}_1^t, \gamma_1^t)$	$\hat{\mathbf{z}}_1^t = g_{z1}(\mathbf{R}_1^t, \tau_1^t)$
$v_1^t = \frac{1}{\beta n}\sum_{i=1}^n \partial_{T_i} g_{x1}(\mathbf{T}_1^t, \gamma_1^t)$	$c_1^t = \frac{1}{\beta m}\sum_{\mu=1}^m \partial_{R_\mu} g_{z1}(\mathbf{R}_1^t, \tau_1^t)$
$\mathbf{T}_2^t = \frac{1}{v_1^t}\hat{\mathbf{x}}_1^t - \mathbf{T}_1^t$	$\mathbf{R}_2^t = \frac{1}{c_1^t}\hat{z}_1^t - \mathbf{R}_1^t$
$\gamma_2^t = \frac{1}{v_1^t} - \gamma_1^t$	$\tau_2^t = \frac{1}{c_1^t} - \tau_1^t$
(Estimation of $\boldsymbol{x}$)	*(Estimation of $\boldsymbol{z}$)*
$\hat{\mathbf{x}}_2^t = g_{x2}(\mathbf{T}_2^t, \mathbf{R}_2^t, \gamma_2^t, \tau_2^t)$	$\hat{\mathbf{z}}_2^t = g_{z2}(\mathbf{T}_2^t, \mathbf{R}_2^t, \gamma_2^t, \tau_2^t)$
$v_2^t = \left\langle \frac{1}{\tau_2^t\lambda + \gamma_2^t} \right\rangle_\nu$	$c_2^t = \frac{1}{\alpha}\left\langle \frac{\lambda}{\tau_2^t\lambda + \gamma_2^t} \right\rangle_\nu$
$\mathbf{T}_1^{t+1} = \frac{1}{v_2^t}\hat{\mathbf{x}}_2^t - \mathbf{T}_2^t$	$\mathbf{R}_1^{t+1} = \frac{1}{c_2^t}\hat{\mathbf{z}}_2^t - \mathbf{R}_2^t$
$\gamma_1^{t+1} = \frac{1}{v_2^t} - \gamma_2^t$	$\tau_1^{t+1} = \frac{1}{c_2^t} - \tau_2^t$

return $\hat{\boldsymbol{x}}_1^T$;

Armed with this equivalent representation of our problem, we can then apply an *adaptive interpolation* strategy to prove the replica-symmetric formula of Proposition 6.3. The strategy is very similar to the detailed proof we presented in Chapter 4: we design a simpler interpolating estimation problem whose free energy is precisely the one of Proposition 6.3, and that is parameterized by the two parameters $q, \hat{q}$. These parameters are then well-chosen along the interpolation path, which allows to deduce the replica-symmetric formula. Since this proof is very similar to the one of the committee machine presented in Chapter 4 we do not reproduce it in this thesis, but the interested reader will find it in detail in [MLKZ20].

6.2.2 Algorithmic point of view: the G-VAMP algorithm

The majority of algorithms developed to solve phase retrieval are based either on semi-definite programming relaxations [CLS15a, Wal18, GS18] or on more direct non-convex optimization procedures, e.g. Wirtinger flow [CLS15b]. The class of *approximate message-passing* algorithms has also been quite successful for specific instances of phase retrieval [SR14, MV21]. Leveraging our usual toolbox, we describe here the most generic algorithm of this class, that allows to tackle any right-rotationally invariant matrix $\mathbf{\Phi}$. It is denoted G-VAMP, and its general form was first written in [SRF16]. Recall that we already studied G-VAMP in relation with the TAP free entropy in Chapter 2. As for many similar problems, G-VAMP is the best-known polynomial time algorithm for this problem in terms of achieved MSE. It makes use of the SVD decomposition of $\mathbf{\Phi}$, that we write as $\mathbf{\Phi}/\sqrt{n} = \mathbf{USV}^\dagger$. The full iterations of the algorithm are detailed in Algorithm 6. We used some auxiliary functions, defined below:

$$g_{x1}(\mathbf{T}, \gamma)_i \equiv \mathbb{E}_{P_0(\gamma, -T_i)}[x], \qquad g_{x2}(\mathbf{T}, \mathbf{R}, \gamma, \tau) \equiv \frac{\mathbf{T}}{\gamma} + \mathbf{V}\mathbf{S}^\intercal\left(\frac{\gamma}{\tau} + \mathbf{S}\mathbf{S}^\intercal\right)^{-1}\left(\frac{\mathbf{U}^\dagger\mathbf{R}}{\tau} - \frac{\mathbf{S}\mathbf{V}^\dagger\mathbf{T}}{\gamma}\right),$$
$$g_{z1}(\mathbf{R}, \tau)_\mu \equiv \mathbb{E}_{P_{\text{out}}\left(y_\mu, \frac{R_\mu}{\tau}, \frac{1}{\tau}\right)}[z], \quad g_{z2}(\mathbf{T}, \mathbf{R}, \gamma, \tau) \equiv \mathbf{USV}^\dagger g_{x2}(\mathbf{T}, \mathbf{R}, \gamma, \tau).$$

$P_0(\gamma, \lambda)$ is the probability distribution with density proportional to $P_0(x)e^{-\frac{\beta\gamma}{2}|x|^2 - \beta\lambda_i \cdot x}$, and $P_{\text{out}}(y_\mu, \omega_\mu, b)$ the one with density proportional to $P_{\text{out}}(y_\mu|z)e^{-\frac{\beta|z-\omega_\mu|^2}{2b}}$.

Statistical and algorithmic performance – Conjecture 6.1 and Theorem 6.2 show that the global maximum of the potential in eq. (6.4) describes the performance of the statistically optimal estimator $\hat{\mathbf{X}}_{\text{opt}}$. Moreover eq. (6.4) also contains rich information about the algorithmic aspects of this problem. Indeed, as for other message-passing algorithms analyzed in the previous chapters, it has been shown that the asymptotic performance of G-VAMP corresponds to the MSE achieved by running gradient descent on the potential in eq. (6.4) from the trivial initial condition $q_x = q_z = 0$ [SRF16]. This allows to derive the thresholds characterizing the statistical and algorithmic limitations of signal estimation solely from eq. (6.4).

State Evolution — The extrema of the potential in eq. (6.4) can be characterized by the solutions of the following *State Evolution* (SE) equations, obtained by looking at the zero-gradient points:

$$\begin{cases} q_x = \mathbb{E}_\xi \mathcal{Z}_0 |f_0|^2, & q_z = \frac{1}{\hat{Q}_z+\hat{q}_z}\big[\frac{\hat{q}_z}{\hat{Q}_z} + \mathbb{E}_\xi \int \mathrm{d}y\ \mathcal{Z}_{\text{out}} |f_{\text{out}}|^2\big], & \text{(6.10a)} \\ \hat{q}_x = \dfrac{q_x}{\rho(\rho - q_x)} - \gamma_x, & \hat{q}_z = \frac{q_z}{Q_z(Q_z-q_z)} - \gamma_z, & \text{(6.10b)} \\ \rho - q_x = \Big\langle \dfrac{1}{\rho^{-1} + \gamma_x + \lambda\gamma_z} \Big\rangle_\nu, & \alpha(Q_z - q_z) = \Big\langle \frac{\lambda}{\rho^{-1}+\gamma_x+\lambda\gamma_z} \Big\rangle_\nu. & \text{(6.10c)} \end{cases}$$

where $f_0(b,a) = \partial_b \ln \mathcal{Z}_0(b,a)$ and $f_{\text{out}}(y;\omega,v) = \partial_\omega \ln \mathcal{Z}_{\text{out}}(y;\omega,v)$ are evaluated at $(b,a) = (\sqrt{\hat{q}_x}\xi, \hat{q}_x)$ and $(\omega,v) = (\sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi, \frac{1}{\hat{Q}_z+\hat{q}_z})$ respectively. Note in particular that the two equations in eq. (6.10c) have to be solved over (γ_x, γ_z) in order to be iterated.

6.3 Weak and perfect recovery transitions

Notation – We adopt the subscript IT for the thresholds related to the Bayes-optimal estimator and Algo for the G-VAMP ones.

6.3.1 Weak recovery: beating a random guess

A natural question to ask is: what is the minimum sample complexity $\alpha_{\text{WR,Algo}} \geq 0$ such that for all $\alpha \geq \alpha_{\text{WR,Algo}}$ we can algorithmically reconstruct $\mathbf{X}^\star$ better than a trivial random draw from the known signal distribution P_0? Also known as the *algorithmic weak-recovery* threshold, $\alpha_{\text{WR,Algo}}$ can also be characterized in terms of the MSE achieved by G-VAMP:

$$\alpha_{\text{WR,Algo}} \equiv \arg\min_{\alpha \geq 0} \{\text{MSE}_{\text{GVAMP}}(\alpha) < \rho\}.$$

Note that we already encountered such weak-recovery transitions (in terms of noise level) in the spiked matrix model of Chapter 5. Since the algorithmic performance is characterized by precisely maximizing eq. (6.4) starting from the trivial point, the threshold $\alpha_{\text{WR,Algo}}$ can be analytically computed from a local stability analysis of this point. On a more general note, an important algorithmic question is to characterize the class of polynomial-time algorithms that can achieve weak recovery directly above the optimal threshold. Beyond G-VAMP, in Section 6.4 we will design simple spectral methods that achieve this feat.

Existence and location of the weak-recovery threshold – It is easy to verify that the state evolution equations (6.10) admit a trivial fixed point in which $q_x = q_z = \hat{q}_x = \hat{q}_z = \gamma_x = \gamma_z = 0$ when P_0 and P_{out} are *symmetric* in the sense of Definition 6.1. When it exists, the trivial extremizer $q_x = q_z = 0$ correspond to either a linearly stable or unstable trivial fixed point of the state evolution equations. The weak-recovery threshold can therefore be determined by looking at the Jacobian around the trivial fixed point. The details of the stability analysis are

given in Appendix D.5.1. The result is that a linear instability of the trivial fixed point appears at $\alpha = \alpha_{\text{WR,Algo}}$ satisfying the equation:

$$\alpha_{\text{WR,Algo}} = \frac{\langle\lambda\rangle_\nu^2}{\langle\lambda^2\rangle_\nu}\left(1 + \left[\int_{\mathbb{R}} \mathrm{d}y \frac{\left|\int_{\mathbb{K}} \mathcal{D}_\beta z\ (|z|^2-1)\ P_{\text{out}}(y|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha_{\text{WR,Algo}}}}z)\right|^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z\ P_{\text{out}}(y|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha_{\text{WR,Algo}}}}z)}\right]^{-1}\right). \tag{6.11}$$

Note that the integrand and the averages $\langle\cdot\rangle_\nu$ depend on $\alpha_{\text{WR,Algo}}$, so that this is an implicit equation on $\alpha_{\text{WR,Algo}}$. Eq. (6.11) is the most generic formula for the weak recovery threshold for any rotationally-invariant data matrix $\mathbf{\Phi}$ and phase retrieval channel P_{out}. As emphasized in the following examples, it generalizes in particular several previously known formulas for different channels and random matrix ensembles.

Gaussian matrix – For a Gaussian i.i.d. matrix, $\langle\lambda\rangle_\nu = \alpha$ and $\langle\lambda^2\rangle_\nu = \alpha^2 + \alpha$, so that

$$\alpha_{\text{WR,Algo}} = \left[\int_{\mathbb{R}} \mathrm{d}y \frac{|\int_{\mathbb{K}} \mathcal{D}_\beta z\ (|z|^2-1)\ P_{\text{out}}(y|\sqrt{\rho}z)|^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z\ P_{\text{out}}(y|\sqrt{\rho}z)}\right]^{-1}, \tag{6.12}$$

a result which was previously derived in [MM19] in the real and complex cases.

Noiseless phase retrieval – In noiseless phase retrieval one has $P_{\text{out}}(y|z) = \delta(y - |z|^2)$. In particular one can easily check that this implies:

$$\alpha_{\text{WR,Algo}} = \left(1 + \frac{\beta}{2}\right)\frac{\langle\lambda\rangle_\nu^2}{\langle\lambda^2\rangle_\nu}. \tag{6.13}$$

This last formula allows to retrieve and generalize many results previously derived in the literature. For instance, for a Gaussian i.i.d. matrix, we find $\alpha_{\text{WR,Algo}} = \beta/2$, which was derived in [MM19, LAL19]. For an orthogonal or unitary column matrix, $\alpha_{\text{WR,Algo}} = 1 + (\beta/2)$, which was already known for $\beta = 2$[MM19] (but not for $\beta = 1$). For the product of p i.i.d. Gaussian matrices with sizes $k_0, \cdots, k_p$, with $k_0 = m$ and $k_p = n$, and $\gamma_l \equiv n/k_l$ for $0 \leq l < p$, we have $\alpha_{\text{WR,Algo}} = (\beta/2)[1 + \sum_{l=1}^{p} \gamma_l]^{-1}$, which generalizes the previously-known real case [ALB+20]. Eq. (6.13) encapsulates all these results and goes beyond by considering an arbitrary spectrum for the sensing matrix, while eq. (6.11) also considers arbitrary channels P_{out}.

The weak-recovery IT transition – We only considered the *algorithmic* weak-recovery threshold. Extending our analysis to the *information-theoretic* threshold $\alpha_{\text{WR,IT}}$ is an interesting open direction, which requires understanding the appearance of a global maximum in the replica-symmetric potential of eq. (6.4), but not necessarily continuously from the $q_x = q_z = 0$ solution. At the moment, we are not able to carry such an analysis.

6.3.2 Perfect recovery for Gaussian signals in noiseless phase retrieval

We consider here noiseless phase retrieval $P_{\text{out}}(y|z) = \delta(y - |z|^2)$ and a prior $P_0 = \mathcal{N}_\beta(0, 1)$[7].

For high number of samples $\alpha \gg 1$, we expect the MMSE to plateau at a minimum achievable reconstruction error $\text{MMSE}_0 \equiv \inf_\alpha \text{MMSE}(\alpha)$, which is a function of the statistics of $\mathbf{\Phi}$. We thus look for the information-theoretical *full-recovery* threshold $\alpha_{\text{FR,IT}}$ as the smallest sample complexity such that MMSE_0 is attained. In Appendix D.5.2 we show that the full-recovery can be *perfect* ($\text{MMSE}_0 = 0$) or *partial* ($\text{MMSE}_0 > 0$) depending on the rank of $\mathbf{\Phi}$ and that

$$\alpha_{\text{FR,IT}} \equiv \beta(1 - \nu(\{0\})). \tag{6.14}$$

[7] We can take $\rho = 1$, as the scaling is irrelevant under a noiseless channel.

Informally, $\nu(\{0\})$, the fraction of zeros in the spectrum of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$, is the fraction of the signal "lost" by the sensing matrix. The effect of rank deficiency is illustrated in Fig. 6.4-left, with $\mathbf{\Phi}$ given by a product of two Gaussian matrices. We emphasize that $\alpha_{\text{FR,IT}}$ is in general not well-defined for an arbitrary channel, which is why we restricted here to the noiseless case. Note that in [BCMN14] it was conjectured that for column-unitary matrices, $\alpha = 4$ measurements were necessary to recover an *arbitrary* signal $\mathbf{X}^\star$. Here we have shown that a *typical* $\mathbf{X}^\star$ can be perfectly recovered as soon as $\alpha = 2$: such possible gaps between worst-case and typical performances are important to keep in mind in a large part of this thesis.

6.3.3 Surprising consequences and open questions

We list here some interesting (and often surprising) consequences of our analysis of the transitions. Since our rigorous results concern a subclass of orthogonally invariant matrices, proving and/or interpreting these statements more generally is an interesting future direction.

- One sees from eq. (6.11) that maximizing $\alpha_{\text{WR,Algo}}$ implies maximizing $\langle\lambda\rangle_\nu^2/\langle\lambda^2\rangle_\nu$. The highest ratio is reached when ν is a delta distribution: for any symmetric channel and prior the ensemble that maximizes $\alpha_{\text{WR,Algo}}$ is thus the one of uniformly-sampled column-orthogonal ($\beta = 1$) or column-unitary ($\beta = 2$) matrices. Conversely, $\alpha_{\text{WR,Algo}}$ can be made arbitrarily small using a product of many Gaussian matrices, both in the real and complex cases.

- In complex noiseless phase retrieval the information-theoretic weak-recovery threshold for column-unitary matrices is located at $\alpha_{\text{WR,IT}} = 2$ [MDX+21]. Our results (Table 6.1) imply that this corresponds to an "all-or-nothing" transition located precisely at $\alpha = 2$. Moreover, our characterization of $\alpha_{\text{WR,Algo}}$ and $\alpha_{\text{FR,IT}}$ shows that for any complex matrix $\alpha_{\text{WR,Algo}} = 2\langle\lambda\rangle_\nu^2/\langle\lambda^2\rangle_\nu \leq \alpha_{\text{FR,IT}} = 2(1-\nu(\{0\}))$, with the equality only being attained for ν a delta distribution. Uniformly sampled column-unitary matrices are thus the only right-unitarily invariant complex matrices which present an "all-or-nothing" transition in complex noiseless phase retrieval (for a Gaussian prior). To the best of our knowledge, this is a first establishment of such a transition in a "dense" problem (as opposed to a sparse setting [GZ17, RXZ19]).

- Consider again noiseless phase retrieval with Gaussian prior. For real orthogonal matrices, one has $\alpha_{\text{WR,Algo}} - \alpha_{\text{FR,IT}} = 1.5 - 1 > 0$: information-theoretic perfect recovery is achieved before algorithmic weak recovery! Since $\alpha_{\text{WR,Algo}}$ is a smooth function of ν, we expect that the inequality holds for many real random matrix ensembles. However, in the complex case, by our previous point, $\alpha_{\text{WR,Algo}} \leq \alpha_{\text{FR,IT}}$ for all matrices: the gap thus only occurs in the real setting.

6.4 Efficient algorithms: constructing optimal spectral methods

6.4.1 Universality of the optimal method

In general, the different optimization methods used to solve phase retrieval (cf. Section 6.2.2) require an "informed" initialization $\hat{\mathbf{X}}$, i.e. that is positively correlated with the signal $\mathbf{X}^*$. The privileged class of algorithms to obtain such initializations in a computationally cheap manner is *spectral methods*, i.e. estimates given by the principal eigenvector of an appropriate matrix constructed from the sensing matrix $\mathbf{\Phi}$ and the observations $\{y_\mu\}$.

In most previous approaches [MM19, LAL19, LL20, MDX+21], the design of spectral methods for the phase retrieval problem was restricted to matrices of the type:

$$\mathbf{M}(\mathcal{T}) \equiv \frac{1}{n}\sum_{\mu=1}^{m} \mathcal{T}(y_\mu)\overline{\Phi_{\mu i}}\Phi_{\mu j}. \tag{6.15}$$

These matrices are functions of a (bounded) *preprocessing* function $\mathcal{T}$. It was previously shown for Gaussian i.i.d. matrices $\mathbf{\Phi}$ [LAL19, LL20] and for random column-unitary matrices $\mathbf{\Phi}$ [DBMM20, MDX+21] that the optimal transition and reconstruction errors in the class of spectral methods described by eq. (6.15) is attained by the following functions:

$$\mathcal{T}^*_{\text{Gaussian}}(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho)}{1+\rho\partial_\omega g_{\text{out}}(y_\mu, 0, \rho)}, \qquad \mathcal{T}^*_{\text{Unitary}}(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho/\alpha)}{1+\frac{\rho}{\alpha}\partial_\omega g_{\text{out}}(y_\mu, 0, \rho/\alpha)}. \tag{6.16}$$

In eq. (6.16) we introduced the function g_{out}, defined as[8]:

$$g_{\text{out}}(y_\mu, \omega, \sigma^2) \equiv \frac{1}{\sigma^2}\frac{\int_{\mathbb{K}} \mathrm{d}x\; e^{-\frac{\beta}{2\sigma^2}|x-\omega|^2}\;(x-\omega)\;P_{\text{out}}(y_\mu|x)}{\int_{\mathbb{K}} \mathrm{d}x\; e^{-\frac{\beta}{2\sigma^2}|x-\omega|^2}\;P_{\text{out}}(y_\mu|x)}. \tag{6.17}$$

The main result of this section is a conjecture that generalizes the above results and gives the optimal spectral method for any phase retrieval problem of the type of eq. (6.2) with right-rotation invariant $\mathbf{\Phi}$[9].

Conjecture 6.5 (*Optimal spectral method in phase retrieval*)

For any right-rotationally (or unitarily) invariant matrix $\mathbf{\Phi}$ (cf. Model R), the optimal (in terms of both weak-recovery transition and achieved reconstruction error) spectral method belongs to the class of eq. (6.15), and is attained by:

$$\mathcal{T}^*(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1+\frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}.$$

Before detailing further our results, let us explicit two important consequences of Conjecture 6.5:

- Note that one can always assume the global scaling $\mathrm{Tr}[\mathbf{\Phi}^\dagger\mathbf{\Phi}]/n^2 \to \langle\lambda\rangle_\nu = \alpha$, as it can be absorbed into the channel P_{out}[10]. The optimal spectral method (in terms of weak-recovery threshold and achieved correlation) is then given by $\mathcal{T}^*(y) = \partial_\omega g_{\text{out}}(y_\mu, 0, \rho)/(1+\rho\partial_\omega g_{\text{out}}(y_\mu, 0, \rho))$. Remarkably, this optimal function *does not depend on the spectrum of the sensing matrix* $\mathbf{\Phi}$*, nor on the sampling ratio* α. The universality of the method is striking when one compares it to the replica result of Conjecture 6.1, which is heavily dependent on the spectrum of $\mathbf{\Phi}$ and on the sampling ratio α. Universality also has deep consequences for phase retrieval practitioners: when using a spectral initialization, she/he does not have to take into account the details of the correlations in $\mathbf{\Phi}$ to construct an optimal spectral method! Although this universality requires in theory right-rotation invariance, this assumption can be loosened as numerically explored in Section 6.5.

- Conjecture 6.5 claims optimality of our method among *all spectral methods* that one can construct from the data $\mathbf{\Phi}$ and the observations $\{y_\mu\}$. It turns out that this optimal method belongs to the class of eq. (6.15), but our derivation is fully constructive and did not assume

[8] In the complex case, $\partial_\omega g_{\text{out}}$ is the "Wirtinger" derivative $\partial_z f(z) \equiv (\partial_x f(z) - i\partial_y f(z))/2$.

[9] Note that Conjecture 6.5 is compatible with the results of eq. (6.16). Indeed, for Gaussian i.i.d. matrices, one has $\langle\lambda\rangle_\nu = \alpha$, while for random column-unitary matrices, $\langle\lambda\rangle_\nu = 1$.

[10] This scaling is chosen to match the one of Gaussian i.i.d. matrices.

anything on the form of the spectral method. This is an important improvement with respect to the previous analyses we mentioned, which always assumed the method to be in the class of eq. (6.15). In this sense, this chapter also confirms the validity of this restriction.

Let us now briefly outline the remaining of this section. Its main purpose is to derive and unify three different approaches to construct optimal spectral methods for the phase retrieval problem. The first one, based on linearizing G-VAMP (Algorithm 6) is studied in Section 6.4.2. In Section 6.4.3 we consider a second approach, based on the *Bethe Hessian* and intimately connected with the TAP picture we derived in Chapter 2. A remarkable property of these two spectral methods is that their derivation is *automated*. Nevertheless, as we show in Section 6.4.4, for any channel distribution and sensing matrix $\mathbf{\Phi}$, the Bethe Hessian method coincides exactly with a third approach, which consists in simply generalizing a spectral method that has been proven to be optimal for Gaussian [LAL19] and unitary [DMM20] sensing matrices, see eq. (6.16). We moreover relate the performance of these three different approaches, and show that they allow to conjecture the optimal spectral method, summarized in Conjecture 6.5.

6.4.2 Linearized vector approximate message passing

This first approach arises as a linearization of G-VAMP (Algorithm 6). It is similar in essence to the technique we used in Chapter 5 to develop spectral methods that leveraged the structure of the signal. Here we apply this method to real and complex phase retrieval with a right-rotationally invariant sensing matrix.

The trivial fixed point

In Algorithm 6, one can use the Nishimori identity to derive the following relation (see for instance eq. (107) of [KKM+16]):

$$\frac{1}{m}\sum_{\mu=1}^{m}\mathbb{E}_{P_{\text{out}}\left(y_\mu,\frac{(R_1^t)_\mu}{\tau_1^t},\frac{1}{\tau_1^t}\right)}\left[\left|z-\frac{(R_1^t)_\mu}{\tau_1^t}\right|^2\right]=\frac{1}{\tau_1^t}. \tag{6.18}$$

Informally, eq. (6.18) expresses that the estimated variance of $\mathbf{z}$, defined as τ_1^t, is equal to the mean square difference between $\mathbf{z}$ and the estimation of $\mathbf{z}$ (being $\mathbf{R}_1^t/\tau_1^t$) under the estimated posterior. Using eq. (6.18) along with the symmetry assumptions of Def. 6.1, it is easy to see that Algorithm 6 admits the following fixed point, that we call "trivial" as it is completely uninformative:

$$\begin{cases} \gamma_1=0, & \gamma_2=\rho^{-1}, & v_1=\rho, & v_2=\rho \\ \hat{\mathbf{x}}_1=\mathbf{T}_1=0, & \hat{\mathbf{x}}_2=\mathbf{T}_2=0, & \tau_1=\alpha/(\rho\langle\lambda\rangle_\nu), & \tau_2=0 \\ c_1=\rho\langle\lambda\rangle_\nu/\alpha, & c_2=\rho\langle\lambda\rangle_\nu/\alpha, & \hat{\mathbf{z}}_1=\mathbf{R}_1=0, & \hat{\mathbf{z}}_2=\mathbf{R}_2=0. \end{cases} \tag{6.19}$$

Linearization around the fixed point

We can now linearize Algorithm 6 around the fixed point given by eq. (6.19). One can easily check (see [MKLZ21] for details) that the first order variations of all the variances and inverse variances parameters are negligible. This will greatly simplify our linearization around the trivial fixed point, as we can focus solely on the vector parameters. For clarity, we restrict here to the real case $\beta=1$, while the publication [MKLZ21] also considered the complex case. We write the linearization of Algorithm 6 as (all derivatives are taken at the fixed point of eq. (6.19)):

$$\begin{aligned} &\delta\hat{\mathbf{x}}_1^t=\nabla_{\mathbf{T}}g_{x1}\delta\mathbf{T}_1^t, && \delta\hat{\mathbf{z}}_1^t=\nabla_{\mathbf{R}}g_{z1}\delta\mathbf{R}_1^t, && \delta\mathbf{T}_2^t=\tfrac{1}{\rho}\delta\hat{\mathbf{x}}_1^t-\delta\mathbf{T}_1^t, \\ &\delta\hat{\mathbf{x}}_2^t=\nabla_{\mathbf{T}}g_{x2}\delta\mathbf{T}_2^t+\nabla_{\mathbf{R}}g_{x2}\delta\mathbf{R}_2^t, && && \delta\hat{\mathbf{z}}_2^t=\nabla_{\mathbf{T}}g_{z2}\delta\mathbf{T}_2^t+\nabla_{\mathbf{R}}g_{z2}\delta\mathbf{R}_2^t, \\ &\delta\mathbf{R}_2^t=\tfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_1^t-\delta\mathbf{R}_1^t, && \delta\mathbf{T}_1^{t+1}=\tfrac{1}{\rho}\delta\hat{\mathbf{x}}_2^t-\delta\mathbf{T}_2^t, && \delta\mathbf{R}_1^{t+1}=\tfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_2^t-\delta\mathbf{R}_2^t. \end{aligned} \tag{6.20}$$

The derivatives of the auxiliary functions of G-VAMP at the trivial fixed point are:

$$\begin{cases} \partial_{T_j}[(g_{x1})_i] = \rho\,\delta_{ij}, & \partial_{T_j}[(g_{x2})_i] = \rho\,\delta_{ij}, & \partial_{R_\nu}[(g_{z1})_\mu] = \delta_{\mu\nu}\mathbb{E}_{P_{\text{out}}\left(y_\mu, 0, \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\right)}[z^2], \\ \partial_{R_\mu}[(g_{x2})_i] = \rho\frac{(\mathbf{\Phi}^\dagger)_{i\mu}}{\sqrt{n}}, & \partial_{T_i}[(g_{z2})_\mu] = \rho\,\frac{\Phi_{\mu i}}{\sqrt{n}}, & \partial_{R_\nu}[(g_{z2})_\mu] = \rho\,\frac{(\mathbf{\Phi}\mathbf{\Phi}^\dagger)_{\mu\nu}}{n}. \end{cases} \tag{6.21}$$

Plugging eq. (6.21) in eq. (6.20) yields, with $v(y_\mu) \equiv \mathbb{E}_{P_{\text{out}}\left(y_\mu, 0, \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\right)}[z^2]$:

$$\begin{cases} \delta\hat{\mathbf{x}}_1^t = \rho\delta\mathbf{T}_1^t, & \delta\hat{\mathbf{z}}_1^t = \text{Diag}(\{v(y_\mu)\})\delta\mathbf{R}_1^t, & \delta\mathbf{T}_2^t = \frac{1}{\rho}\delta\hat{\mathbf{x}}_1^t - \delta\mathbf{T}_1^t, \\ \delta\mathbf{R}_2^t = \frac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_1^t - \delta\mathbf{R}_1^t, & \delta\hat{\mathbf{x}}_2^t = \rho\delta\mathbf{T}_2^t + \rho\frac{\mathbf{\Phi}^\dagger}{\sqrt{n}}\delta\mathbf{R}_2^t, & \delta\hat{\mathbf{z}}_2^t = \rho\frac{\mathbf{\Phi}}{\sqrt{n}}\delta\mathbf{T}_2^t + \rho\frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n}\delta\mathbf{R}_2^t, \\ \delta\mathbf{T}_1^{t+1} = \frac{1}{\rho}\delta\hat{\mathbf{x}}_2^t - \delta\mathbf{T}_2^t, & \delta\mathbf{R}_1^{t+1} = \frac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_2^t - \delta\mathbf{R}_2^t. \end{cases} \tag{6.22}$$

In particular, these equations imply $\delta\mathbf{T}_2^t = 0$. The equations can then simply be closed on $\delta\mathbf{R}_1^t$:

$$\delta\mathbf{R}_1^{t+1} = \Big(\frac{\alpha}{\langle\lambda\rangle_\nu}\frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n} - \mathrm{I}_m\Big)\Big[\frac{\alpha}{\rho\langle\lambda\rangle_\nu}\text{Diag}(\{v(y_\mu)\}) - \mathrm{I}_m\Big]\delta\mathbf{R}_1^t. \tag{6.23}$$

In the complex case, one obtains the same linearized equation. Interestingly, $v(y_\mu)$ can be linked to the function $\partial_\omega g_{\text{out}}$ of eq. (6.17): $\partial_\omega g_{\text{out}}(y_\mu, 0, \sigma^2) = -\sigma^{-2} + \sigma^{-4}v(y_\mu)$.

The LAMP spectral method

The Linearized-AMP (LAMP) spectral method is based on eq. (6.23), and consists in taking the dominant eigenvalue and corresponding eigenvector of the $m \times m$ matrix:

$$\mathbf{M}^{(\text{LAMP})} \equiv \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\Big(\frac{\alpha}{\langle\lambda\rangle_\nu}\frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n} - \mathrm{I}_m\Big)\text{Diag}(\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)).$$

Note that $\mathbf{M}^{(\text{LAMP})}$ is not a Hermitian matrix, so "dominant" eigenvalue means here eigenvalue of largest real part. If $\hat{\mathbf{u}}$ is the eigenvector of $\mathbf{M}^{(\text{LAMP})}$ associated to this dominant eigenvalue, then one can construct a corresponding estimate $\hat{\mathbf{x}}$ using the relations of eq. (6.22), as:

$$\hat{\mathbf{x}} \equiv \frac{\mathbf{\Phi}^\dagger\big[\frac{\alpha}{\rho\langle\lambda\rangle_\nu}\text{Diag}(\{v(y_\mu)\}) - \mathrm{I}_m\big]\hat{\mathbf{u}}}{\Big\|\mathbf{\Phi}^\dagger\big[\frac{\alpha}{\rho\langle\lambda\rangle_\nu}\text{Diag}(\{v(y_\mu)\}) - \mathrm{I}_m\big]\hat{\mathbf{u}}\Big\|}\sqrt{n\rho}. \tag{6.24}$$

Surprisingly, and as we will see in more details in the following, this spectral method achieves the optimal weak-recovery threshold but only sub-optimal performance compared to $\mathbf{M}(\mathcal{T}^*)$! There is, however, a way to recover the optimal performance from $\mathbf{M}^{(\text{LAMP})}$ by considering an eigenvalue "hidden" inside the bulk, as we will see in Proposition 6.6.

6.4.3 The Bethe Hessian: TAP revisited

Our second approach leverages the Thouless-Anderson-Palmer (TAP) formalism of statistical physics [TAP77], and in particular the results of Chapter 2.

Let us first recall a few facts on the TAP formalism. It consists in studying a modified posterior distribution with respect to eq. (6.3), which is tilted so that the first and second moments of all x_i are fixed. These moments become then variables of the *TAP free energy* associated with this modified posterior distribution. When weak recovery of the signal is impossible, this free energy possesses a global minimum in the completely uninformative point in which the estimator is the vector $\mathbf{m} = 0$. On the other hand, when weak recovery is possible, the optimal estimator corresponds to the global minimum of the TAP free energy with $\mathbf{m} \neq 0$. However as we will

see the point $\mathbf{m} = 0$ always remains a stationary point of the TAP free energy. Moreover, a spectral method used for initializing a non-convex optimization algorithm can be based solely on the observations (i.e. on $\mathbf{\Phi}$ and $\{y_\mu\}$), and therefore can not exploit any physical information other than the one present in the uninformative point. When this point is locally stable, we therefore expect all polynomial-time algorithms not to be able to achieve weak recovery. On the other hand, linear instability of the $\mathbf{m} = 0$ point implies that there should exist a minimum of the TAP free entropy with positive correlation with the signal. With this picture in mind, it is natural to conjecture that the optimal spectral estimator is the dominant unstable direction of the uninformative fixed point, i.e. the smallest eigenvalue of the Hessian at the uninformative fixed point, also denoted *Bethe Hessian*[11].

The TAP free entropy

We denote $f_{\text{TAP}}(\mathbf{Y}, \mathbf{\Phi}, \mathbf{m}, \sigma)$ the TAP free entropy (or negative free energy) of the model of eq. (6.2). Of particular interest are its maxima, corresponding to *pure states* in the statistical physics language. Recall that we showed in Chapter 2 that these pure states are in exact correspondence with the fixed points of the G-VAMP algorithm. Let us now use the results of Chapter 2 to compute f_{TAP}, more specifically eq. (2.39), which yields up to $\mathcal{O}_n(1)$ terms:

$$\begin{aligned} f_{\text{TAP}}(\mathbf{m}) = \sup_{\sigma\geq 0} \sup_{\substack{\mathbf{g}\in\mathbb{K}^m \\ r\geq 0}} \operatorname*{extr}_{\substack{\boldsymbol{\omega}\in\mathbb{K}^m \\ b\geq 0}} \operatorname*{extr}_{\substack{\boldsymbol{\lambda}\in\mathbb{K}^n \\ \gamma\geq 0}} \Big[& \frac{\beta}{n}\sum_{i=1}^n \lambda_i \cdot m_i + \frac{\beta\gamma}{2n}(n\sigma^2 + \sum_{i=1}^n |m_i|^2) - \frac{\beta}{n}\sum_{\mu=1}^m \omega_\mu \cdot g_\mu \\ & - \frac{\beta b}{2n}(\sum_{\mu=1}^m |g_\mu|^2 - \alpha n r) + \frac{1}{n}\sum_{i=1}^n \ln \int_{\mathbb{K}} P_0(\mathrm{d}x) e^{-\frac{\beta\gamma}{2}|x|^2 - \beta\lambda_i \cdot x} \\ & + \frac{\alpha}{m}\sum_{\mu=1}^m \ln \int_{\mathbb{K}} \frac{\mathrm{d}h}{(\frac{2\pi b}{\beta})^{\beta/2}} P_{\text{out}}(y_\mu|h) e^{-\frac{\beta|h-\omega_\mu|^2}{2b}} + \frac{\beta}{n}\sum_{i=1}^n\sum_{\mu=1}^m g_\mu \cdot (\frac{\Phi_{\mu i}}{\sqrt{n}} m_i) + \beta F(\sigma^2, r) \Big]. \end{aligned} \tag{6.25}$$

The function F is defined as:

$$F(x,y) \equiv \inf_{\zeta_x,\zeta_y>0} \Big[\frac{\zeta_x x}{2} + \frac{\alpha\zeta_y y}{2} - \frac{\alpha-1}{2}\ln\zeta_y - \frac{1}{2}\langle \ln(\zeta_x\zeta_y + \lambda)\rangle_\nu \Big] - \frac{1}{2}\ln x - \frac{\alpha}{2}\ln y - \frac{1+\alpha}{2}.$$

One can write the saddle-point equations associated to eq. (6.25), called the *TAP equations*:

$$\begin{aligned} & m_i = \mathbb{E}_{P_0(\gamma,\lambda_i)}[x], && \sigma^2 = \tfrac{1}{n}\sum_{i=1}^n \mathbb{E}_{P_0(\gamma,\lambda_i)}[|x-m_i|^2], \\ & g_\mu = g_{\text{out}}(y_\mu, \omega_\mu, b), && r = \tfrac{1}{m}\sum_{\mu=1}^m \Big\{ |g_\mu|^2 + \tfrac{1}{b} - \mathbb{E}_{P_{\text{out}}(y_\mu,\omega_\mu,b)}\Big[\big|\tfrac{h-\omega_\mu}{b}\big|^2 \Big] \Big\}, \\ & \omega_\mu + b g_\mu = \sum_{i=1}^n \tfrac{\Phi_{\mu i}}{\sqrt{n}} m_i, && b = -\tfrac{2}{\alpha}\partial_r F(\sigma^2, r), \quad \gamma = -2\partial_{\sigma^2} F(\sigma^2, r). \end{aligned} \tag{6.26}$$

The trivial stationary point

It is easy to see that the TAP equations (6.26) admit a trivial fixed point at $\mathbf{m} = 0$ (corresponding to a stationary point of f_{TAP}). At this point, the parameters are $\sigma^2 = \rho$, $\mathbf{g} = \boldsymbol{\omega} = \boldsymbol{\lambda} = 0$, $\gamma = r = 0$, $b = \rho\langle\lambda\rangle_\nu/\alpha$. The derivation of the fixed point also uses the behavior of $F(\sigma^2, r)$ at small r, computed in Appendix D.6.1:

$$F(\sigma^2, r) = -\frac{\langle\lambda\rangle_\nu r\sigma^2}{2} + \frac{\sigma^4 r^2}{4\alpha}[\alpha\langle\lambda^2\rangle_\nu - (1+\alpha)\langle\lambda\rangle_\nu^2] + \sigma^6 r^3 G(r\sigma^2), \tag{6.27}$$

with $G(x)$ a continuous bounded function in $x = 0$.

[11] The Bethe Hessian has also been used as a spectral method e.g. in the context of community detection [SKZ14]

The spectral method

As we argued, a natural way to design the optimal spectral method for this inference problem is to consider the Hessian of $-f_{\rm TAP}$ at this trivial fixed point, as we expect a descending informative direction to appear in its spectrum at the weak recovery threshold. Computing the Hessian from eq. (6.25) is quite lengthy, but poses no technical difficulties. For this reason we do not reproduce its derivation here (it can be found in [MKLZ21]). In the end this procedure leads to consider the $n \times n$ matrix:

$$\mathbf{M}^{(\rm TAP)} \equiv -\frac{n}{\beta}\nabla^2 f_{\rm TAP}(0) = -\frac{1}{\rho}\mathrm{I}_n + \frac{1}{n}\sum_{\mu=1}^{m} \frac{\partial_\omega g_{\rm out}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\rm out}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)} \overline{\Phi}_{\mu i}\Phi_{\mu j}. \tag{6.28}$$

6.4.4 Unifying the approaches

We detail here our main analytical results concerning the spectral methods we just derived.

The optimal spectral method and the Bethe Hessian

Very importantly, as opposed to previous approaches, our derivation is *constructive*: we start from the fully-explicit expression of the TAP free entropy given in eq. (6.25) and simply compute its Hessian at the trivial fixed point. As we argued, we expect from the statistical physics literature that the optimal spectral method will be given by the largest eigenvalue (and associated eigenvector) of this Hessian, given in eq. (6.28). Importantly, this implies that the optimal spectral method that can be built from the data $\mathbf{\Phi}$ and the observations $\{y_\mu\}$ belongs to the class of methods given by eq. (6.15). Our conjecture therefore also gives weight to many previous analysis of spectral methods for phase retrieval, which only studied spectral methods of the type of eq. (6.15) [MM19, LAL19, LL20, MDX+21].

Relating linearized-AMP and the Bethe Hessian

Our derivation of $\mathbf{M}^{(\rm LAMP)}$ is *constructive* as well, and in this sense fundamentally differs from the L-AMP algorithms designed in [MDX+21] to assess the performance of other spectral methods. We can moreover relate the eigenpairs of our two methods:

Proposition 6.6 (*Relating LAMP to the Bethe Hessian*)

Without loss of generality, we assume $\langle\lambda\rangle_\nu = \alpha$. Let $z_\mu \equiv \partial_\omega g_{\rm out}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)$, and $(\lambda_{\rm LAMP}, \mathbf{v})$ be an eigenpair of $\mathbf{M}^{(\rm LAMP)}$. Assume that $\lambda_{\rm LAMP} + \rho z_\mu \neq 0$ for all $\mu = 1, \cdots, m$. Then $\mathbf{\Phi}^\dagger \mathrm{Diag}(z_\mu)\mathbf{v} \neq 0$, and we let $\hat{\mathbf{x}} \propto \mathbf{\Phi}^\dagger \mathrm{Diag}(z_\mu)\mathbf{v}$ with $\|\hat{\mathbf{x}}\|^2 = n$. Moreover:

$$\Big\{\frac{1}{m}\sum_{\mu=1}^{m} \frac{\rho z_\mu}{\lambda_{\rm LAMP} + \rho z_\mu}\mathbf{\Phi}_\mu \mathbf{\Phi}_\mu^\dagger\Big\}\hat{\mathbf{x}} = \hat{\mathbf{x}}.$$

Conversely, let $\mathbf{x}$ be an eigenvector of $\mathbf{M}^{(\rm TAP)}$ with norm $\|\mathbf{x}\|^2 = n$, with associated eigenvalue $\lambda_{\rm TAP}$. We define $\mathbf{u} \equiv \mathrm{Diag}[(1+\rho z_\mu)^{-1}]\mathbf{\Phi}\mathbf{x}/\sqrt{n}$. Then one has:

$$\mathbf{M}^{(\rm LAMP)}\mathbf{u} = \mathbf{u} + \rho\lambda_{\rm TAP}\mathrm{Diag}(1 + \rho\partial_\omega g_{\rm out}(y_\mu, 0, \rho))\mathbf{u}.$$

Moreover, if $\lambda_{\rm TAP} = 0$, eq. (6.24) applied to $\mathbf{u}$ yields the same performance as the TAP estimator.

This proposition is proven in Appendix D.6.2. By considering $\lambda_{\rm LAMP} = 1$ and $\lambda_{\rm TAP} = 0$, one immediately deduces two important consequences of Proposition 6.6:

- The appearance of an unstable direction, in the spectrum of $\mathbf{M}^{(\mathrm{TAP})}$ (i.e. a positive eigenvalue) and of $\mathbf{M}^{(\mathrm{LAMP})}$ (i.e. an eigenvalue with real part greater than 1), occurs at a common threshold, the *weak-recovery* threshold, given by eq. (6.11).
- An eigenvalue 0 appears in the spectrum of $\mathbf{M}^{(\mathrm{TAP})}$ if and only if an eigenvalue 1 appears in the spectrum of $\mathbf{M}^{(\mathrm{LAMP})}$. These two eigenvalues therefore correspond to *marginal stability* of the linear dynamics. Moreover, the two estimators associated to these eigenvalues are identical, i.e. $\mathbf{M}^{(\mathrm{LAMP})}$ *contains the optimal estimator*. However, this estimator is different from the dominant eigenvector of $\mathbf{M}^{(\mathrm{LAMP})}$, which reaches only suboptimal performance as we will see in Section 6.5.

6.5 Statistical and algorithmic analysis: numerical experiments

This section is devoted to a numerical investigation of all the theoretical claims of the previous sections. In Section 6.5.1 we solve the state evolution equations (6.10) for different real and complex ensembles of sensing matrix $\mathbf{\Phi}$, and compare it to numerical simulations of G-VAMP. This allows to uncover the existence or absence of hardness depending on the structure of the sensing matrix. In the following Section 6.5.2 we give the performance of the spectral methods we derived in Section 6.4, for noiseless and Poisson-noise observations. Finally, in Section 6.5.3 we apply both message-passing and spectral algorithms to the recovery of real images, illustrating the relevance of our analysis for practical phase retrieval.

We also show in all these cases that the algorithms we present perform very well even by allowing more structure in the sensing matrix than assumed in Model R, by considering for example randomly subsampled DFT, Hadamard or DCT matrices[12].

6.5.1 Optimal algorithms and computational gaps

While our results hold for any phase retrieval problem (in the sense of Def. 6.1), we focus for the analysis of computational gaps on noiseless phase retrieval. We fix $P_{\mathrm{out}}(y|z) = \delta(y - |z|^2)$ and take $P_0 = \mathcal{N}_\beta(0,1)$. We can indeed consider $\rho = 1$, as the scaling is irrelevant under a noiseless channel. The numerical code used to generate all simulations of the G-VAMP algorithm is available in a Github repository [MLKZ20].

Real case

In Fig. 6.2, we illustrate the case of real Gaussian and real column-orthogonal sensing matrix $\mathbf{\Phi}$, the latter not having been investigated previously in the literature. We compute the MMSE by solving the State Evolution equations starting from an *informed* solution (close to full recovery). The minimal mean-squared error achievable with the G-VAMP algorithm is computed using the State Evolution equations starting from the *uninformed* $q_z = 0$ solution. We compare these predictions with numerical simulations of the G-VAMP algorithm on Gaussian matrices and uniformly sampled orthogonal matrices, as well as randomly subsampled Hadamard matrices. The simulations are in very good agreement with the prediction, and our results on Hadamard matrices suggest that the curves of Fig. 6.5-right are valid for more general ensembles than uniformly sampled orthogonal matrices, and that one can allow some controlled structure in the matrix without harming the performance of the algorithm.

[12] The universality of linearized approximate message passing algorithms for a Gaussian prior and different ensembles of column-orthogonal matrices was analyzed recently in [DB20].

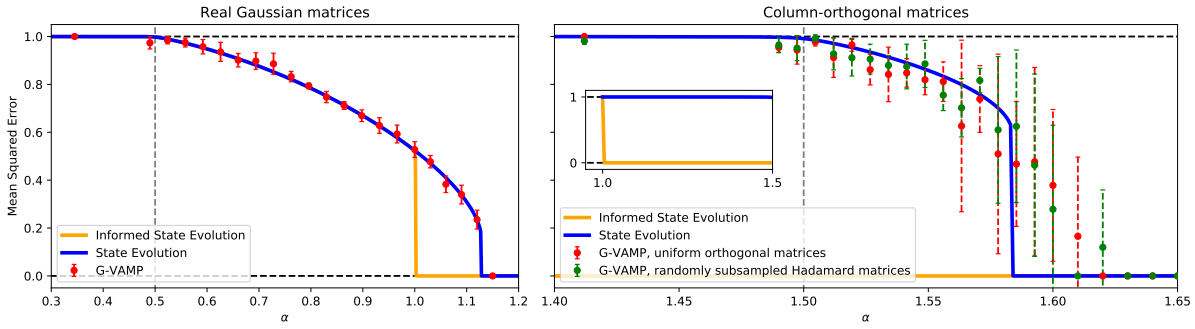


FIGURE 6.2: Comparison of the Bayes-optimal MSE and the G-VAMP algorithm, for an i.i.d. real Gaussian (left) and a column-orthogonal (right) sensing matrix $\mathbf{\Phi}$ (i.e. $\mathbf{\Phi}^\intercal\mathbf{\Phi}/n = \mathrm{I}_n$), with a real Gaussian prior. Dots correspond to finite size simulations of G-VAMP (the mean and std are taken over 5 instances, with $n = 8000$ in the Gaussian case and $m = 8192$ in the orthogonal case), while full lines are obtained from the state evolution. The vertical grey dashed lines denote the algorithmic weak recovery threshold $\alpha_{\mathrm{WR,Algo}}$. Note the presence of a statistical-to-algorithmic gap in both ensembles, and that for column-orthogonal matrices $\alpha_{\mathrm{WR,Algo}} > \alpha_{\mathrm{FR,IT}}$.

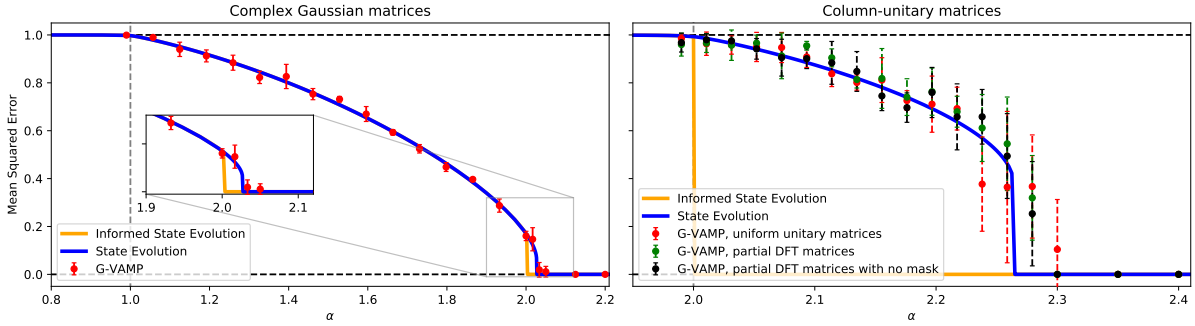


FIGURE 6.3: We show quantities identical to Fig. 6.2, but for an i.i.d. complex Gaussian (left) and a column-unitary (right) sensing matrix $\mathbf{\Phi}$ (i.e. $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n = \mathrm{I}_n$), with a complex Gaussian prior. We used $n = 5000$ in the simulations of G-VAMP.

Complex case

Previous works on complex signals $\mathbf{X}^* \in \mathbb{C}^n$ have mainly focused on the study of the weak recovery threshold α_{WR}, which was located for Gaussian matrices [MM19, LAL19] and uniformly sampled column-unitary matrices [MP17, DBMM20]. We begin by extending the aforementioned results by identifying the full recovery threshold $\alpha_{\mathrm{FR,IT}}$ in these cases, and comparing the performance of the G-VAMP algorithm to the Bayes-optimal solution. Fig. 6.3 illustrates our results for these two ensembles. The algorithmic (i.e. G-VAMP) full-recovery threshold $\alpha_{\mathrm{FR,Algo}}$ is found numerically from the state evolution equations and is in good agreement with finite size simulations. The existence of a statistical-to-algorithmic gap $\Delta = \alpha_{\mathrm{FR,Algo}} - \alpha_{\mathrm{FR,IT}} \geq 0$ reflects the hardness of phase retrieval in the real and complex case. However, it is interesting to note that even though full-recovery in the complex case requires more data than in the real case, the size of the statistical-to-algorithmic gap in the complex ensembles is smaller than in their real counterparts. Again our conclusions transfer to *partial DFT matrices*, which were introduced in [MYP14, MDX+21], and are column-unitary matrices obtained from the usual DFT matrices. Namely, there are defined for $m \geq n$ as $\mathbf{\Phi}/\sqrt{n} = \mathbf{FSP}$, with $\mathbf{F} \in \mathbb{C}^{m\times m}$ a DFT matrix, $\mathbf{S} \in \mathbb{R}^{m\times n}$ containing n columns (randomly taken) of the identity matrix I_m, and $\mathbf{P}$ a diagonal of random phases.

In Fig. 6.4 we analyze the case of a product of two i.i.d. standard Gaussian matrices $\mathbf{\Phi} =$

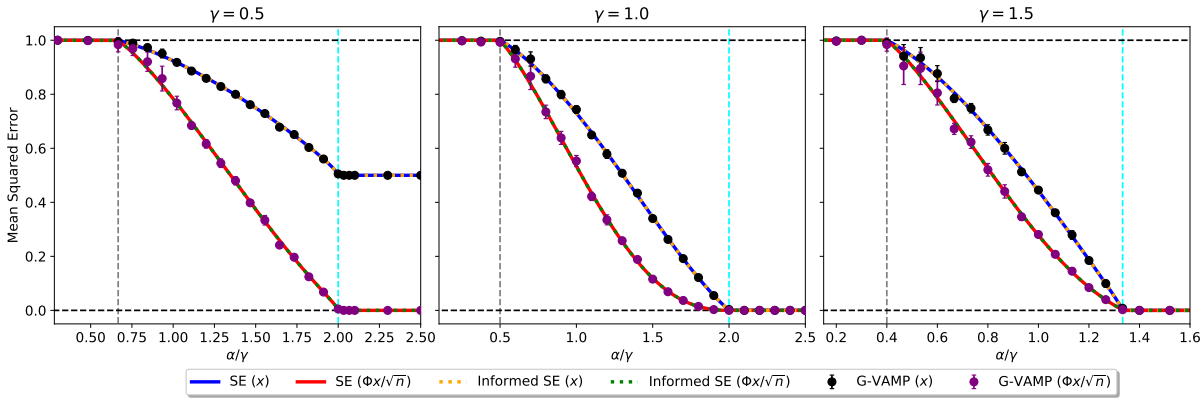


FIGURE 6.4: Mean squared error as a function of the measurement rate α, for a sensing matrix $\mathbf{\Phi} = \mathbf{W}_1\mathbf{W}_2$ a product of two complex i.i.d. standard Gaussian matrices $\mathbf{W}_1 \in \mathbb{C}^{m\times p}$, $\mathbf{W}_2 \in \mathbb{C}^{p\times n}$ with aspect ratios $\gamma = p/n \in \{0.5, 1.0, 1.5\}$. Red curves denote the recovery on $\mathbf{\Phi}\mathbf{X}^\star/\sqrt{n}$ and blue curves on $\mathbf{X}^\star$. Cyan dashed lines denote the full reconstruction threshold $\alpha_{\text{FR,IT}}$. The G-VAMP experiments were performed with $n = 5000$, and the mean and std are taken over 5 instances.

$\mathbf{W}_1\mathbf{W}_2$, with $\mathbf{W}_1 \in \mathbb{C}^{m\times p}$ and $\mathbf{W}_2 \in \mathbb{C}^{p\times n}$ for different aspect ratios $\gamma \equiv p/n$. We can identify the presence of a threshold $\alpha_{\text{WR,Algo}} = \gamma/(1+\gamma)$ (computed in Section 6.3.1) that delimits the possibility of weak recovery both information-theoretically and in polynomial time. The information-theoretic full-recovery is achieved at $\alpha_{\text{FR,IT}} = \min(2, 2\gamma)$, in agreement with eq. (6.14). Note that the full recovery algorithmic threshold is very close to the information-theoretic one, and precisely equal for $\gamma = 1$, although the gap is too small to be visible in the left and right parts of Fig. 6.4. Therefore, the performance of G-VAMP is exactly given by the Bayes-optimal estimator, apart for $\gamma \neq 1$ in a very small range $(\alpha_{\text{FR,IT}}, \alpha_{\text{FR,Algo}})$, whose size is of order 10^{-3} for $\gamma \in \{0.5, 1.5\}$. As $\gamma \to \infty$, one recovers the statistical-to-algorithmic gap present in the complex Gaussian case, which is again very small (around 0.027, cf Table 6.1). Although this hard phase is very small, our results show its existence for all $\gamma \neq 1$.

6.5.2 Spectral methods: cheap and efficient

In this section, we numerically assess our predictions and compare the performances of the spectral methods on various problems. The numerical code used to generate all figures related to the spectral methods is available in a Github repository [MKLZ20].

Another spectral method – In the figures, we sometimes consider another spectral method, denoted $\mathbf{M}^{(\text{MM})}$. It is obtained by naively considering the preprocessing function of [MM19], which was shown to achieve the optimal transition for Gaussian sensing matrices. More precisely, we have (assuming $\rho = 1$ and $\langle\lambda\rangle_\nu = \alpha$):

$$\mathcal{T}_{\text{MM}}(y) \equiv \frac{\partial_\omega g_{\text{out}}(y, 0, 1)}{\sqrt{\frac{2\alpha}{\beta}} + \partial_\omega g_{\text{out}}(y, 0, 1)}.$$

In particular note that at $\alpha = \beta/2$, we have $\mathcal{T}_{\text{MM}} = \mathcal{T}^*$, so that $\mathcal{T}_{\text{MM}}$ indeed achieves the optimal weak-recovery transition for Gaussian matrices, for which $\alpha_{\text{WR,Algo}} = \beta/2$.

Performance of the spectral methods – We show the performance of the spectral methods to recover a random signal in three different cases, that we briefly describe:

- In Fig. 6.5, we consider noiseless real phase retrieval (i.e. sign retrieval), with uniformly sampled column-unitary sensing matrices. We also show that our conclusions transfer to randomly subsampled Hadamard matrices, validating the conclusions of [DB20].

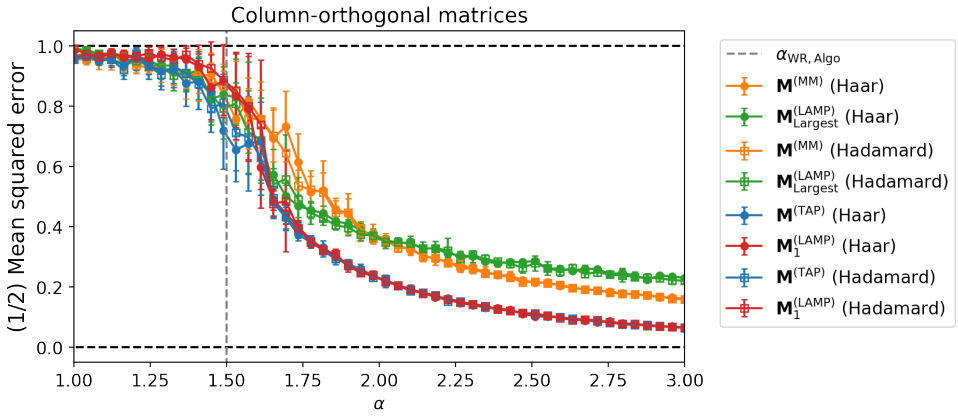


FIGURE 6.5: MSE achieved by our spectral methods and a naive version of the spectral method of [MM19] for real column-orthogonal sensing matrices and a noiseless channel. We give the performance on uniformly sampled column-orthogonal matrices as well as randomly subsampled Hadamard matrices. The simulations were done using $m = 8192$, and the error bars are taken over 10 instances.

- In Fig. 6.6a we consider noiseless real phase retrieval when the sensing matrix is a product of two Gaussian i.i.d. matrices. This setup can for instance be interpreted as Gaussian phase retrieval in which the signal is drawn from a known generative prior, similarly to what we analyzed in Chapter 5. Importantly, it is not covered by any previous analysis of the spectral methods, emphasizing the generality of Conjecture 6.5.

- In Fig. 6.6b, we compare our results in noiseless and noisy settings. More precisely, we consider complex phase retrieval with a Gaussian sensing matrix, and either a noiseless channel or a Poisson observation channel with intensity $\Lambda > 0$:

$$P_{\text{out}}(y|z) = e^{-\Lambda|z|^2} \sum_{k=0}^{\infty} \delta(y-k) \frac{\Lambda^k |z|^{2k}}{k!}.$$

  This latter channel is particularly relevant for optical applications, in which the detector can be modeled as being affected by a Poisson noise. In both cases, we find that all our conclusions on the optimality of the $\mathbf{M}^{(\text{TAP})}$, and on the link between $\mathbf{M}^{(\text{LAMP})}$ and $\mathbf{M}^{(\text{TAP})}$, still hold.

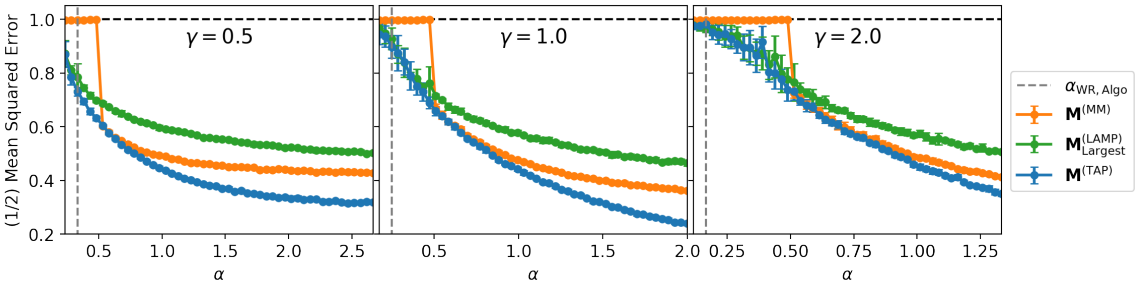
Transition phenomena in the spectra – We illustrate the weak-recovery transition in the spectra of the different methods. Precisely, we confirm the following claims of Section 6.4.4:

- Both $\mathbf{M}^{(\text{LAMP})}$ and $\mathbf{M}^{(\text{TAP})}$ have a dominant eigenvalue that detaches from the bulk for $\alpha > \alpha_{\text{WR,Algo}}$, given by eq. (6.11).

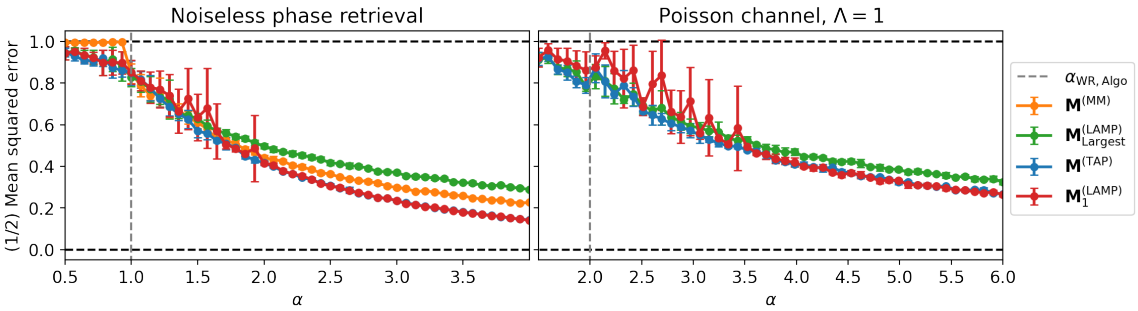
- In the regime in which weak-recovery is possible, the largest eigenvalue of $\mathbf{M}^{(\text{TAP})}$ approaches 0 as $n \to \infty$. The associated eigenvector achieves optimal correlation with the signal (among spectral methods) as $n \to \infty$.

- $\mathbf{M}^{(\text{LAMP})}$ gives *two* estimators that are positively correlated with the signal for $\alpha > \alpha_{\text{WR,Algo}}$. The first one corresponds to its largest eigenvalue in real part, and achieves worse correlation than the largest eigenvector of $\mathbf{M}^{(\text{TAP})}$. The second one corresponds to an eigenvalue inside the bulk (but isolated from the other eigenvalues) that approaches 1 as $n \to \infty$, and achieves the same optimal performance as the estimator given by $\mathbf{M}^{(\text{TAP})}$.

We verify these claims for different values of α, below and above the weak-recovery threshold, in complex Gaussian phase retrieval with Poisson-noise, in Fig. 6.7. This analysis is extended in



(A) Product of two real i.i.d. Gaussian sensing matrices with a size ratio $\gamma \in \{0.5, 1.0, 2.0\}$. The simulations were done using $m = 10000$, and error bars are taken over 10 instances.



(B) Complex Gaussian matrix, in noiseless phase retrieval and in Poisson-noise phase retrieval with $\Lambda = 1$. The simulations were done using $m = 10000$ (noiseless case), 12000 (Poisson case), and the error bars are taken over 10 (noiseless case), 5 (Poisson case) instances.

FIGURE 6.6: MSE achieved by the different spectral methods in two different settings.

[MKLZ21], which contains similar results for other observation channels and sensing matrices.

Computational cost of the spectral methods – When weak recovery is possible the largest eigenvalue of $\mathbf{M}^{(\mathrm{TAP})}$ concentrates on 0 as we noticed. However, the spectrum of $\mathbf{M}^{(\mathrm{TAP})}$ also contains many very large negative eigenvalues. In practice, we use an inverse iteration method to quickly estimate the associated eigenvector. We use a similar approach for $\mathbf{M}^{(\mathrm{LAMP})}$, using inverse iterations to estimate the eigenvector with eigenvalue 1, and usual power iterations for the largest eigenvalue.

A puzzling open question – As we already said, the optimal estimator is always associated with *marginal stability*, both in $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$. A clear understanding of this marginal stability is still lacking. Moreover, the dominant eigenvector of the matrix $\mathbf{M}^{(\mathrm{LAMP})}$ is associated to an *unstable* direction, thus dominating the dynamics of the linearized-AMP. However its achieved correlation is smaller than the one achieved by the marginally stable, optimal, eigenvector. We also noticed that the eigenvectors of $\mathbf{M}^{(\mathrm{TAP})}$ *do not contain any information about this suboptimal estimator*[13]. This blindness of $\mathbf{M}^{(\mathrm{TAP})}$ to the principal eigenvector of $\mathbf{M}^{(\mathrm{LAMP})}$ is very puzzling from a theoretical point of view. Indeed, as we showed in Chapter 2, the stationary limit of G-VAMP (Algorithm 6) is in exact correspondence with the stationary point equations of the TAP free entropy. One would therefore expect the two spectral methods $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$ to contain the same physical information on the system. Moreover, the different qualitative behaviors of the two dynamics (instability of $\mathbf{M}^{(\mathrm{LAMP})}$ a opposed to marginal stability of $\mathbf{M}^{(\mathrm{TAP})}$) only deepens this puzzle.

[13] In particular, this is an important distinction between our L-AMP constructive derivation and the L-AMP algorithms of [MDX+21], which are *designed* to match the spectral methods of the type $\mathbf{M}(\mathcal{T})$: in the latter, it was shown that the L-AMP estimator always matched the one of the spectral method.

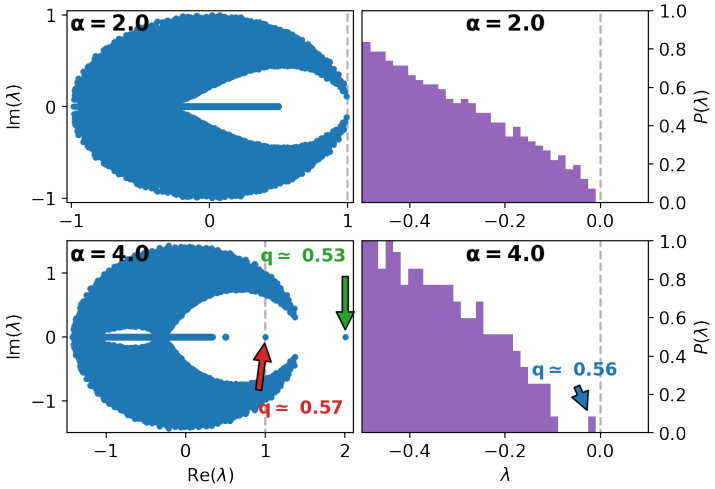


FIGURE 6.7: Transition in the spectra of $\mathbf{M}^{(\mathrm{LAMP})}$ (left) and $\mathbf{M}^{(\mathrm{TAP})}$ (right) for a complex Gaussian $\mathbf{\Phi}$ and a Poisson channel with $\Lambda = 1$. For $\alpha > \alpha_{\mathrm{WR,Algo}} = 2$, we indicate the approximate overlap q corresponding to the the relevant eigenvalues.

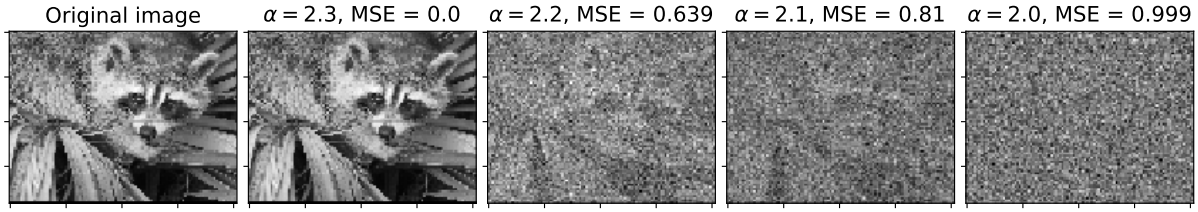


FIGURE 6.8: Performance of the G-VAMP algorithm for noiseless phase retrieval. We wish to recover a 77x102 image (on the left), and we use a complex Gaussian prior to infer the signal. The data matrix $\mathbf{\Phi}$ is a randomly subsampled DFT matrix.

6.5.3 Real image reconstruction

Importantly, while the knowledge of the distribution of the true signal is required for our theoretical analysis, the G-VAMP algorithm and our spectral methods are also well-defined beyond this scope, e.g. they can be used to infer natural images with Fourier matrices. Using a Gaussian prior to infer the image can then actually be seen as the minimal assumption on the underlying signal, as it amounts to simply fix its norm: in particular, our theory can thus predict the performance of G-VAMP for any signal, structured or not.

Performance of G-VAMP

We conducted a simple experiment on a natural image with a randomly subsampled DFT matrix $\mathbf{\Phi}$, described in Fig. 6.8. Although we are far from a Bayes-optimal setting, the achieved MSE is very close to values of Fig. 6.3, for all values of α. In particular, we achieve perfect recovery for $\alpha \geq 2.3$, just above $\alpha_{\mathrm{FR,Algo}} \simeq 2.27$ which was derived for random unitary matrices, i.i.d. data and in the Bayes-optimal setting!

Spectral initialization on real images

As a final analysis, we numerically investigate our spectral methods for the reconstruction of a natural image. For comparability, we consider the image of *The Birth of Venus* already used in [MM19, MDX+21]. Although this signal is not i.i.d., we will see that all the conclusions that we numerically investigated in Section 6.5.2 transfer to this case. We consider a noiseless

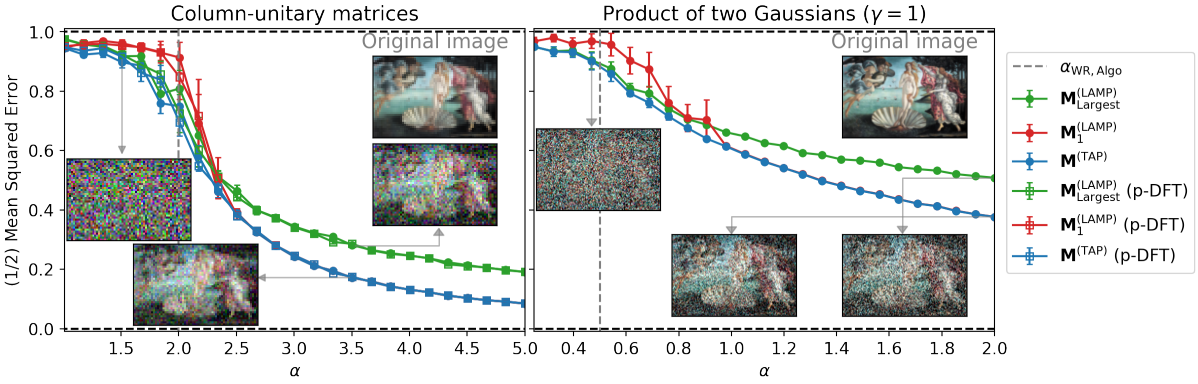


FIGURE 6.9: MSE achieved by the different spectral methods for the recovery of a natural image in noiseless phase retrieval. We consider column-unitary matrices $\mathbf{\Phi}$ (both uniformly sampled and partial DFT matrices, left) and the product of two complex Gaussian matrices with aspect ratio $\gamma = 1$ (right). We reduced each dimension of the original 1280×820 image by a factor 20 (left) or 10 (right), and we average the MSE over 5 instances and the 3 RGB channels.

phase retrieval channel and different sensing matrices $\mathbf{\Phi}$: multiple ensembles of column-unitary matrices (which partly reproduces the analysis of [MDX+21]) and a product of two complex Gaussian matrices with aspect ratio $\gamma = 1$. In Fig. 6.9, we give the MSE obtained by the different spectral methods and these two matrix ensembles. We also give examples of the images recovered by the algorithms. Moreover, despite the fact that the signal (and possibly the matrix as well) is structured, we still observe the same transition phenomena in the spectra of $\mathbf{M}^{(\text{TAP})}$ and $\mathbf{M}^{(\text{LAMP})}$. Namely, we still observe that the optimal estimator is associated with marginal stability of both spectral methods, while the dominant eigenvalue of $\mathbf{M}^{(\text{LAMP})}$ is associated to a non-optimal estimator.

Let us finally illustrate how the optimal spectral method of Conjecture 6.5 can be combined with a subsequent local optimization algorithm. We use the spectral estimator as the initialization point to running vanilla gradient descent on the square loss

$$L(\mathbf{x}) \equiv \frac{1}{2m} \sum_{\mu=1}^{m} \left\{ \left| \frac{(\mathbf{\Phi x})_\mu}{\sqrt{n}} \right|^2 - \left| \frac{(\mathbf{\Phi X}^*)_\mu}{\sqrt{n}} \right|^2 \right\}^2 .$$

This allows to already obtain a perfect recovery of the image for $\alpha = 4$, as shown in Fig. 6.10. In [MKLZ21] we describe the obtained MSE curves in more details. In particular, we confirm that combining the gradient descent with the spectral initialization allows to reach perfect recovery at finite α, which is not possible with the "vanilla" spectral methods.

Conclusion of Chapter 6 and perspectives on Part II

In this last well-filled chapter on the statistical physics approach to optimal learning in inference models we considered phase retrieval, one of the flagship models of theoretical computer science, with a wide variety of applications across scientific fields (notably through optical experiments). Our main findings can be briefly summarized as:

- As in Chapters 4 and 5, we leverage the replica method to derive a single-letter formula for the asymptotic free entropy (or mutual information) of the system, which we then prove under assumptions on the structure of the data matrix. Combined with message-passing algorithms, which are known to be optimal among a large class of general first order methods [CMW20], we uncovered the existence or absence of statistical-to-computational gaps in phase retrieval

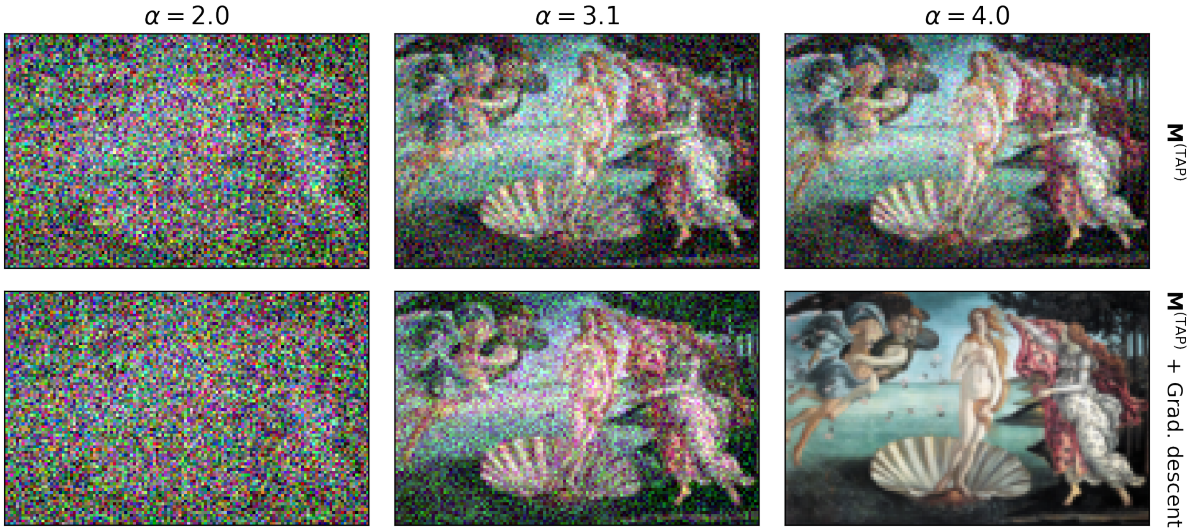


FIGURE 6.10: Reconstruction of a real image in noiseless phase retrieval with partial DFT matrices. We reduce the image size from 1280×820 to 128×82. We compare, for three different values of α, the estimators of $\mathbf{M}(\mathcal{T}^*)$ (top line) and the estimator obtained by running a gradient descent procedure starting from the estimator of $\mathbf{M}(\mathcal{T}^*)$ (bottom line). We recover the 3 RGB channels with independent instances of the sensing matrix.

depending on the structure of the sensing matrix. In particular, for the case of Fourier phase retrieval, which is of particular interest to the practitioner, we uncover a significant hard phase, while the IT performance shows an "all-or-nothing" transition, see Fig. 6.3. We also conducted a precise theoretical analysis of the weak and full recovery thresholds, as a function of the sampling ratio, in the wide class of phase retrieval problems characterized by Def. 6.1. Recall that these results are summarized in Table 6.1.

- We presented an automated derivation of a generic spectral method, which we conjecture to be optimal for the wide class of phase retrieval problems that satisfy Def. 6.1, cf. Conjecture 6.5. Our derivation leverages the *Bethe Hessian* of the TAP free energy computed in Chapter 2, and in particular, we show that our method achieves the optimal weak-recovery threshold, and give numerical evidence for its optimality in terms of MSE.

- Finally, we gave numerical evidence for two generalizations of our results, regarding both optimal learning and the spectral methods: first, the i.i.d. nature of the signal is not required for our results to hold, as we show by recovering real structured images, see Figs. 6.8 and 6.9. Second, one can allow more structure in the data matrix than rotational invariance, as illustrated in Figs. 6.2, 6.3, 6.8 and 6.9. In particular, we can consider *randomly subsampled Fourier matrices* without harming any of our conclusions. Note that this random subsampling is necessary, as we found that our conclusions did not transfer to "vanilla" Fourier matrices.

Proofs of replica formulas for inference: beyond Gaussianity – As we emphasized, the adaptive interpolation method used at several points in Part II, while being highly versatile and applied to a wide range of problems [BM19a, BM19b, BKM+19, Bar19, BR20, BMDK20],[BMMK18], deeply relies on some Gaussianity in the problem, either in the sensing matrix or the prior. Recent works by collaborators have taken a different path, leveraging known results on the asymptotic performance of approximate message-passing algorithms, which allows to indirectly prove the replica predictions [GAK20a, GAK20b]. This has allowed to put on rigorous ground formulas for matrices without any Gaussianity (i.e. the generic rotationally-invariant matrices described by Models S and R). Still these approaches also have limitations,

e.g. requiring convexity of the loss, and an improvement in this regard, or a way to loosen the Gaussianity assumptions in adaptive interpolation, would both be significant advancements in putting the replica predictions in Bayes-optimal problems on rigorous grounds.

Additional references – Let us conclude by naming some references for the reader interested in the ideas behind the whole Part II, and who wishes to know more. First and foremost, [MM09] is an important introduction to the field of statistical physics applied to theoretical computer science. It details many of the concepts that we introduced in this thesis, e.g. message-passing algorithms, the cavity method or the appearance of replica symmetry breaking in inference or constraint satisfaction problems. With a point of view more focused on statistical gaps in inference problems, and close to the one we took in this thesis, [ZK16] gives a very detailed description of how to apply the statistical physics toolbox to learning. The authors focus on important examples particularly relevant to the theoretical computer science community, such as compressed sensing and community detection. In the same vein, [BPW18] reviews some of the statistical physics prediction on the existence of hard phases, and offers a very good introduction to the field for the mathematics and theoretical computer science audience. Finally, closer to the physics perspective, [Gab20] is a recent review on the mean-field approaches for learning in neural networks, and highlights many very recent advances and perspectives. Beyond these reviews, the reader will find many recent published works taking similar approaches (e.g. [BMMK18] by the author), that one can regroup under the broad category of "statistical physics of learning".

Part III

Towards a topological approach to high-dimensional optimization

Chapter 7

The complexity of high-dimensional landscapes

"As the light grew a little he saw to his surprise that what from a distance had seemed wide and featureless flats were in fact all broken and tumbled. Indeed the whole surface of the plains of Gorgoroth was pocked with great holes, as if, while it was still a waste of soft mud, it had been smitten with a shower of bolts and huge slingstones. The largest of these holes were rimmed with ridges of broken rock, and broad fissures ran out from them in all directions."

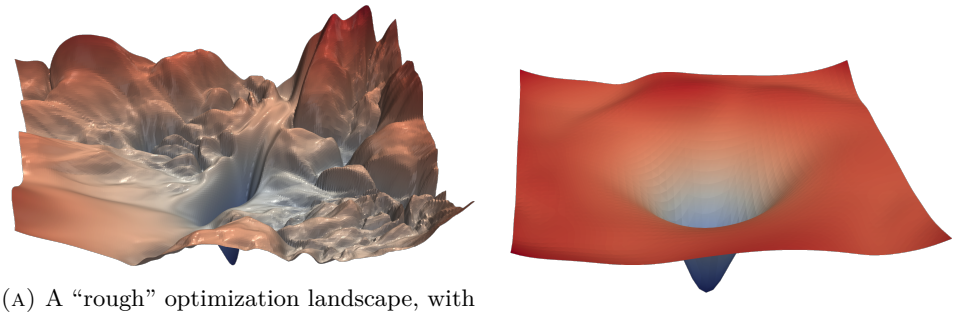
J.R.R. Tolkien, The Lord of the Rings – Book VI (1955).

Disclaimer – In this chapter we present a method to obtain the average and the typical value of the number of critical points of the empirical risk landscape for generalized linear estimation problems and variants. This represents a substantial extension of previous computations which restricted to Gaussian random functions. The new results in this chapter are mainly based on [MBAB20], and the introductory Section 7.1 builds on a presentation given at the Kavli Institute for Theoretical Physics during the winter of 2019 [Mai19]. Some proofs and details are given in Appendix E. On a general note, the majority of this chapter is written in a style closer to the mathematics literature than the other chapters in this thesis. It aims however at being accessible to any reader from the statistical physics community.

7.1 Counting complexity: the Kac-Rice formula

7.1.1 How to "count" the complexity of a landscape?

Characterizing the landscape of the empirical risk is a key issue in several contexts. Indeed, many current machine learning problems (such as the ones we encountered in the previous chapters) are both non-convex and high-dimensional. In these cases the analysis of local optimization algorithms, such as gradient descent and its stochastic variants, represents a very hard feat. In recent years, there has been a series of works that developed a landscape-based approach to tackle this challenge. The key idea is to study the statistical properties of the empirical risk landscape, and to use these findings to obtain results on the performance of algorithms. Without the aim of being exhaustive this research avenue includes analysis of the landscape of neural networks, matrix completion, tensor factorization and tensor principal component analysis (PCA) [Fyo04, FN12, Kaw16, SC16, GLM16, FB17, BNS16, PKCS17, DLT+18, GM17, GJZ17, LK17, LXB19, BAMMN19, RBABC19, SMKUZ19, SMBC+19, BCRT20, BAGJ20]. The majority of these works identifies the region of parameters where the landscape is "easy", i.e. it focuses on the regime where there shouldn't be any bad local minima and it proves that indeed there are none. However, gradient descent and other landscape-based algorithms are used extensively in machine learning, and they are often observed to work even very far from



(A) A "rough" optimization landscape, with a great number of spurious local minima and saddle points.

(B) A smooth landscape, intuitively easier to optimize via local algorithms.

FIGURE 7.1: "Naive" representation of a rough (left) and smooth (right) optimization landscape. Both landscapes have a global minimum with the same value. Pictures taken from [LXT+18].

the region described above where the landscape can be proved mathematically to be "easy". A possible reason is that the bounds obtained rigorously are not tight enough. Another, more interesting, is that the landscape is "hard", i.e. spurious minima are present, but their basins of attraction are small and the dynamics is able to avoid them [SMKUZ19, SMBC+19].

In Chapters 4 and 6, we uncovered large hard phases in the learning of two-layers neural networks and in the phase retrieval problem. This led us to the conclusion that, in the loss landscapes of such models, there must exist a large number of *spurious* local minima (i.e. local minima with loss value far from the global minimum), so that local optimization procedures (as well as all polynomial-time algorithms) fail whenever the system is in a hard phase. The existence of many such local minima is what we will generically describe as the *complexity* of a high-dimensional landscape.

In Fig. 7.1 we describe the intuitive view of complexity of a landscape. Crucially, one must keep in mind that such a picture can be very misleading, as almost all our usual low-dimensional intuitions become incorrect when entering the realm of high dimensions[1].

Counting critical points – In this chapter we develop a general method that allows to study the "hard" regime, where the loss landscape may display a huge number of bad minima. Our aim is to obtain explicit formulas for the number of critical points of the *empirical risk* landscape, and to characterize the Hessian associated to them. For a given problem, this will allow to identify the topological transition where the landscape becomes "easy", and to analyze very precisely the "hard" regime by investigating the topology of its sublevel sets. In recent years, there has been remarkable progress on this subject in the field of spin glasses and probability theory [FSW07, FW07, BD07, ABA13, ABAČ13, Sub17a, Sub17b, SZ17, BAMMN19, RBABC19, BASZ20] (this list being far from exhaustive, and we will refer to many other works adopting similar or complementary approaches in specific paragraphs of this chapter). This line of research has allowed to put on a firm ground results previously obtained in the physics literature [BM80, Kur91, CS95, CGG99, Fyo04], and it has unveiled important relationships with random matrix theory. Its main domain of application has been the study of the landscapes associated to Gaussian random functions, and we will describe one of these analyses in Section 7.1.4. Its extension to tackle non-Gaussian high-dimensional random functions is an open problem — one that is crucial to address in order to characterize the landscape of the empirical risk.

Annealed and quenched complexity – One of the earliest and most important insights of the physics analysis of spin glasses is that for random smooth functions in high dimensions, we should expect the number of critical points to typically be in the scale $e^{\Theta(n)}$ [BM80, Mon95].

[1] For instance, picture the unit sphere $\mathbb{S}^{n-1}$ with a "North pole" $\mathbf{e}_n$. As $n \to \infty$ almost all its mass is concentrated on the equator orthogonal to $\mathbf{e}_n$, which is not intuitive when representing it in 3 dimensions!

Therefore one must be particularly careful as how to consider its asymptotic value, as the number of critical points may be highly fluctuating. In Section 1.1 we introduced two generic asymptotic limits for such random variables, that we denoted as *annealed* and *quenched*, borrowing from the physics jargon. We refer the reader to this section for a more precise description of these two limits. We therefore define the annealed and quenched *complexity* of a random function f_n as:

$$\begin{cases} \Sigma^{\text{annealed}} & \equiv \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\, \text{Crit}(f_n) \\ \Sigma^{\text{quenched}} & \equiv \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \ln\, \text{Crit}(f_n) \end{cases}, \tag{7.1}$$

$\text{Crit}(f_n)$ being the number of critical points of f_n. In the following we will generalize this definition, to count e.g. the local minima, or critical points with given value of the function.

Usually the annealed complexity is easier to compute than its quenched counterpart. However, since $\text{Crit}(f_n)$ is in general exponentially large and strongly fluctuating, the quenched complexity really describes the typical value of the number of critical points, and is truly representative of the typical properties of the landscape, while the annealed complexity might be dominated by rare instances, as e.g. in tensor PCA [BAMMN19, RBABC19]. Annealed and quenched complexities are therefore usually different, with very few exceptions [CS95, Sub17a], one of them being the pure p-spin model that we will see in detail in Section 7.1.4.

Our main tool to count the critical points of random functions will be the *Kac-Rice formula* [Ric44]. In the following Sections 7.1.2 and 7.1.3 we describe its derivation and the intuition behind it, before describing a step-by-step first application to a spin glass model in Section 7.1.4.

7.1.2 The area formula

The Kac-Rice formula is intuitively not derived from any involved probabilistic tool, but rather is a consequence of a purely geometric result, called the *area formula* (itself a consequence of the co-area formula), described in great generality by Federer [Fed59], and stated for instance in [AW09b]. This formula is the generalization of the following non-rigorous intuition: for a smooth function $f : \mathbb{R} \to \mathbb{R}$, and $T \subseteq \mathbb{R}$, denoting $N_f(u, T)$ the number of solutions to the equation $f(t) = u$ with $t \in T$, one would want to write informally:

$$N_f(u, T) = \int_{f(T)} \delta(v - u)\mathrm{d}v = \int_T \delta(f(t) - u)|f'(t)|\mathrm{d}t. \tag{7.2}$$

The area formula generalizes and makes rigorous this intuition, by showing this last equality in the weak sense. We follow here the statement of [AW09b].

Proposition 7.1 (*Area formula, from [AW09b]*)

Let $f : U \to \mathbb{R}^d$ be a $\mathcal{C}^1$ function defined on an open subset U of $\mathbb{R}^d$. Assume that the set of critical values[2] of f has zero Lebesgue measure, and denote $N_f(u, T)$ the number of solutions to the equation $f(t) = u$ with $t \in T$. Then, for any Borel set $T \subseteq \mathbb{R}^d$, and any $g : \mathbb{R}^d \to \mathbb{R}$ continuous and bounded:

$$\int_{\mathbb{R}^d} g(u) N_f(u, T) = \int_T |\det f'(t)|\, g(f(t))\, \mathrm{d}t.$$

Note that in a large part of the theoretical physics literature, the area formula (and subsequently the Kac-Rice formula) is directly written in the form of eq. (7.2).

[2] i.e. points t for which $f'(t)$ is singular.

7.1.3 The Kac-Rice formula

Consider a smooth compact manifold $\mathcal{M}$ of dimension n (a practical example is the n-dimensional unit sphere $\mathbb{S}^{n-1}$), equipped with a Riemannian metric, and an associated volume measure $\mu_{\mathcal{M}}$. We are given a random smooth function $f : \mathcal{M} \to \mathbb{R}$ and we want to use the area formula (Prop. 7.1) to estimate the moments of the number of critical points of f, in order to compute the complexity defined by eq. (7.1).

Given the assumptions of Proposition 7.1, a reasonable hypothesis is to assume that f is almost surely a *Morse* function [MSWW63], i.e. that *all its critical points are non-degenerate*. Since $\mathcal{M}$ is compact, one easily deduces that the number of critical points of a Morse function is finite[3]. For any $k \in \mathbb{N}$ and Borel set $B \subseteq \mathbb{R}$, we define $\mathrm{Crit}_{f,k}(B)$ to be the number of critical points $x \in \mathcal{M}$ of f such that $f(x) \in B$ and such that the index of $\mathrm{Hess}\, f(x)$ (that is the number of strictly negative eigenvalues of the Hessian) is equal to k. The informal area formula of eq. (7.2) applied to $\mathrm{grad}\, f$ reads:

$$\mathrm{Crit}_{f,k}(B) = \int_{\mathcal{M}} \mathrm{d}\mu_{\mathcal{M}}(x) \delta(\mathrm{grad}\, f(x)) |\det \mathrm{Hess}\, f(x)| \mathbb{1}[f(x) \in B, \mathrm{i}(\mathrm{Hess}\, f(x)) = k]$$

Taking the expectation of this equality, one directly obtains the Kac-Rice formula:

Proposition 7.2 (*Kac-Rice formula, informal*)

Let $\mathcal{M}$ be a smooth compact Riemannian manifold of dimension n, with volume measure μ_M. Let $f : \mathcal{M} \to \mathbb{R}$ be a random function that is almost surely Morse, and that satisfies some regularity properties (see the remark below). Denote $\varphi(0)$ the density of $\mathrm{grad}\, f(x)$ with respect to the Lebesgue measure on $\mathbb{R}^{n-1}$, taken at 0. Then:

$$\mathbb{E}\,\mathrm{Crit}_{f,k}(B) = \int_{\mathcal{M}} \mathrm{d}\mu_{\mathcal{M}}(x) \mathbb{E}[|\det \mathrm{Hess}\, f(x)| \mathbb{1}\{f(x) \in B, \mathrm{i}_{\mathrm{Hess}\, f(x)} = k\} | \mathrm{grad}\, f(x) = 0] \varphi(0).$$

Let us make a few important remarks on this formula:

1. The rigorous derivation of the Kac-Rice formula is quite involved, as one has to start from the weak equality of Proposition 7.1 and to use continuity arguments in order to obtain a strong equality at $u = 0$. This proof is detailed e.g. in [AW09b]. This is the first reason for which the formula has mostly been used for Gaussian random fields, since then continuity arguments can be justified using simple hypotheses on the covariance of the random field. The second reason is that in general, conditional expectations of non-Gaussian random variables are intractable, making the Kac-Rice formula effectively useless since one has to know the law of the Hessian conditioned by the gradient being zero. Under many heavy technical conditions, one can however derive rigorous non-Gaussian versions of the Kac-Rice formula (see for instance [AT09, AW09b]). We will apply such sophisticated results in Section 7.3.

2. The Kac-Rice formula transforms a random differential geometry problem into a *random matrix theory* problem. The main difficulty in evaluating the Kac-Rice formula comes from the distribution of the Hessian conditioned by the gradient being zero: even for Gaussian random fields, this is in general a heavily correlated Gaussian random matrix, for which very few results exist.

3. The Kac-Rice formula can be generalized to compute higher moments of $\mathrm{Crit}_{f,k}(B)$ as well (cf. Theorem 6.3 of [AW09b]). Via Morse's theory, it can also be used to compute the moments of the Euler characteristic of the level sets of f, see [ABA13] for an example.

[3] The numbers of critical points of different *indices* (i.e. the number of strictly negative Hessian eigenvalues) of a Morse function are constrained by the topology of $\mathcal{M}$ by the *Morse inequalities* [MSWW63].

7.1.4 The complexity of the pure spherical p-spin model

To conclude our introduction to the Kac-Rice formula, we will describe its application in a flagship model of disordered systems: the (pure) *spherical p-spin*, introduced in [CS92], and that we defined in Model 1.3. This calculation was first performed by physicists [CS95, CGG99, Fyo04, FSW07, FW07], and made rigorous in [ABAČ13] (and subsequent papers). We will follow the derivation of the latter, and we will often refer to it for technicalities: the goal of this section is not to present novel results, but rather to give the reader an intuition of the mechanisms of the Kac-Rice method.

Statement of the problem

Consider $n \geq 1$, $p \geq 3$, and the following function $f_{n,p}$ on the unit sphere $\mathbb{S}^{n-1}$:

$$f_{n,p}(\boldsymbol{\sigma}) \equiv \sum_{1 \leq i_1, \cdots, i_p \leq n} J_{i_1, \cdots i_p} \boldsymbol{\sigma}_{i_1} \cdots \boldsymbol{\sigma}_{i_p} \qquad (\boldsymbol{\sigma} \in \mathbb{S}^{n-1}), \tag{7.3}$$

in which $J_{i_1,\cdots i_p} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. In the physics language, the Hamiltonian $H_{n,p}$ of the pure spherical p-spin defined in Model 1.3 is related to $f_{n,p}$ by $f_{n,p}(\boldsymbol{\sigma}) = n^{-1/2} H_{n,p}(\sqrt{n}\boldsymbol{\sigma})$. For any $B \subseteq \mathbb{R}$, we want to compute the large n limit of the expectation of the number of local minima $\boldsymbol{\sigma}$ of $f_{n,p}$, such that $f_{n,p}(\boldsymbol{\sigma}) \in \sqrt{n}B$, that we denote $\mathrm{Crit}^0_{n,p}(B)$. One can apply the Kac-Rice formula 7.2[4]:

$$\begin{aligned} \mathbb{E}\,\mathrm{Crit}^0_{n,p}(B) = \int_{\mathbb{S}^{n-1}} \mathrm{d}\mu(\boldsymbol{\sigma}) \varphi_{\mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma})}(0) \\ \mathbb{E}\Big[|\det \mathrm{Hess}\, f_{n,p}(\boldsymbol{\sigma})|\, \mathbb{1}\{f_{n,p}(\boldsymbol{\sigma}) \in \sqrt{n}B,\, \mathrm{Hess}\, f_{n,p}(\boldsymbol{\sigma}) \geq 0\} \Big| \mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma}) = 0 \Big], \end{aligned} \tag{7.4}$$

in which μ is the usual surface measure on $\mathbb{S}^{n-1}$. Note that here grad and Hess stand for the *Riemannian* gradient and Hessian on the sphere, while we will denote ∇, ∇^2 the *Euclidean* gradient and Hessian.

The joint distribution of $(f_{n,p}(\boldsymbol{\sigma}), \mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma}), \mathrm{Hess}\, f_{n,p}(\boldsymbol{\sigma}))$

Deriving the joint law of $(f_{n,p}(\boldsymbol{\sigma}), \mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma}), \mathrm{Hess}\, f_{n,p}(\boldsymbol{\sigma}))$ is a necessary step in the Kac-Rice method, as these three random variables appear in the conditional expectation.

We fix $\boldsymbol{\sigma} \in \mathbb{S}^{n-1}$. Note that $(f_{n,p}(\boldsymbol{\sigma}), \mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma}), \mathrm{Hess}\, f_{n,p}(\boldsymbol{\sigma}))$ is a Gaussian centered random variable since f is a Gaussian random field. We thus simply need to compute its correlations to fully characterize the joint distribution. We will naturally identify the tangent space $\mathcal{T}_{\boldsymbol{\sigma}}(\mathbb{S}^{n-1})$ with $\mathbb{R}^{n-1}$. If we denote $P^\perp_{\boldsymbol{\sigma}}$ the orthogonal projector on $\{\boldsymbol{\sigma}\}^\perp$, one has:

$$\begin{aligned} \mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma}) &= P^\perp_{\boldsymbol{\sigma}} \nabla f_{n,p}(\boldsymbol{\sigma}), \\ \mathrm{Hess}\, f_{n,p}(\boldsymbol{\sigma}) &= P^\perp_{\boldsymbol{\sigma}} \nabla^2 f_{n,p}(\boldsymbol{\sigma}) P^\perp_{\boldsymbol{\sigma}} - \langle \boldsymbol{\sigma}, \nabla f_{n,p}(\boldsymbol{\sigma}) \rangle P^\perp_{\boldsymbol{\sigma}}. \end{aligned}$$

For instance one can compute the covariance of the gradient:

$$\mathbb{E}[\mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma}) \mathrm{grad}\, f_{n,p}(\boldsymbol{\sigma})^\intercal] = P^\perp_{\boldsymbol{\sigma}} \mathbb{E}[\nabla f_{n,p}(\boldsymbol{\sigma}) \nabla f_{n,p}(\boldsymbol{\sigma})^\intercal] P^\perp_{\boldsymbol{\sigma}} = p P^\perp_{\boldsymbol{\sigma}}.$$

[4] The proof that $f_{n,p}$ is almost surely a Morse function can be found in [ABAČ13]

Using the same kind of simple algebraic calculations, one obtains the joint distribution:

$$\begin{cases} f_{n,p}(\boldsymbol{\sigma}) & \overset{\mathrm{d}}{=} Z, \\ \operatorname{grad} f_{n,p}(\boldsymbol{\sigma}) & \overset{\mathrm{d}}{=} \sqrt{p}\mathbf{g}, \\ \operatorname{Hess} f_{n,p}(\boldsymbol{\sigma}) & \overset{\mathrm{d}}{=} \sqrt{(n-1)p(p-1)}\mathbf{M}_{n-1} - pZ\,\mathrm{Id}_{n-1}, \end{cases}$$

in which $Z \sim \mathcal{N}(0,1)$, $\mathbf{g} \sim \mathcal{N}(0, \mathrm{I}_{n-1})$, and $\mathbf{M}_{n-1} \sim \mathrm{GOE}(n-1)$ (recall that we defined the GOE in Section 1.5) and the variables $(Z, \mathbf{g}, \mathbf{M}_{n-1})$ are pairwise independent. Let us make the following important remarks:

- The joint distribution of $(f_{n,p}(\boldsymbol{\sigma}), \operatorname{grad} f_{n,p}(\boldsymbol{\sigma}), \operatorname{Hess} f_{n,p}(\boldsymbol{\sigma}))$ is independent of $\boldsymbol{\sigma}$.
- The variables $(f_{n,p}(\boldsymbol{\sigma}), \operatorname{Hess} f_{n,p}(\boldsymbol{\sigma}))$ are independent from $\operatorname{grad} f_{n,p}(\boldsymbol{\sigma})$.
- From the gradient distribution, one easily obtains its density evaluated in 0:

$$\varphi_{\operatorname{grad} f_{n,p}(\boldsymbol{\sigma})}(0) = e^{-\frac{n-1}{2}\ln(2\pi p)}.$$

Recalling that the volume of the unit sphere is $V(\mathbb{S}^{n-1}) = 2\pi^{n/2}/\Gamma(n/2)$, one deduces from the Kac-Rice formula (7.4) and the remarks above:

$$\mathbb{E}\,\mathrm{Crit}^0_{n,p}(B) = \frac{2\pi^{n/2}}{\Gamma(n/2)} e^{\frac{n-1}{2}\ln\frac{(n-1)(p-1)}{2\pi}} \mathbb{E}[|\det \mathbf{H}_{n-1}| \mathbb{1}(\mathbf{H}_{n-1} \geq 0,\, z \in \sqrt{n}B)], \tag{7.5}$$

in which

$$\mathbf{H}_{n-1} \equiv \mathbf{M}_{n-1} - \sqrt{\frac{p}{(n-1)(p-1)}} z \mathrm{I}_n,$$

with $z \sim \mathcal{N}(0,1)$ and $\mathbf{M}_{n-1} \sim \mathrm{GOE}(n-1)$. It is now completely explicit that we reduced our original random differential geometry problem (counting the number of critical points of a random function) to a random matrix theory problem.

Simplification of the formula

It is clear from eq. (7.5) that the following lemma (from [ABAČ13]), which computes the average absolute value of the determinant of shifted GOE matrices, will be crucial.

Lemma 7.3 (*Expectation of the absolute determinant of a shifted GOE matrix*)

Let $G \subseteq \mathbb{R}$ a Borel set, $X \sim \mathcal{N}(0, t^2)$ (for a $t > 0$) and $\mathbf{M}_{n-1} \sim \mathrm{GOE}(n-1)$. Then:

$$\begin{aligned} &\mathbb{E}[|\det(\mathbf{M}_{n-1} - X\mathrm{I}_{n-1})| \mathbb{1}((\mathbf{M}_{n-1} - X\mathrm{I}_{n-1}) \geq 0, X \in G)] \\ &= \frac{2^{\frac{n}{2}}\Gamma(n/2)(n-1)^{-n/2}}{\sqrt{\pi t^2}} \mathbb{E}_{\mathrm{GOE}(n)}\Big[e^{-\frac{n\lambda_0^2}{4}\left(\frac{2}{(n-1)t^2}-1\right)} \mathbb{1}\Big(\lambda_0 \in \sqrt{\frac{n-1}{n}}G\Big)\Big]. \end{aligned}$$

Here λ_0 is the smallest eigenvalue of a random matrix from the $\mathrm{GOE}(n)$ ensemble.

Lemma 7.3 is a result obtained purely from random matrix theory, and is a particular case of Lemma 3.3 of [ABAČ13]. We refer to this work for its proof.

Let us make a very important remark here. While the explicit calculation of Lemma 7.3 is possible because of the GOE structure of the matrix, the appearance of the smallest eigenvalue should not come as a surprise. We indeed wrote the Kac-Rice formula for the number of *local minima* of the function $f_{n,p}$, i.e. critical points with positive Hessian matrix. In order to condition

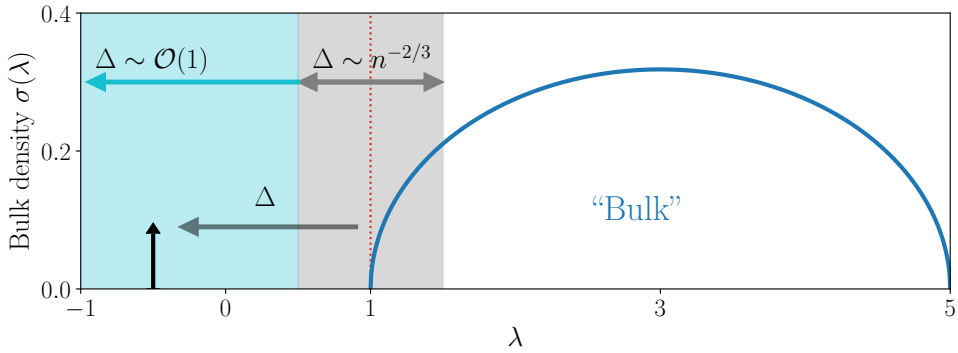


FIGURE 7.2: Generic illustration of a *large deviations* event in which the smallest eigenvalue of a random matrix (here a shifted GOE matrix) is macroscopically far from its expected value (here 1). Δ is the shift of $\lambda_{\max}$ from its expected value, at the left edge of the bulk. We emphasize that while the smallest eigenvalue typically fluctuates in the scale $n^{-2/3}$, as shown in the grey area (these fluctuations are connected to the Tracy-Widom law [TW94]), large deviations instead correspond to *macroscopic* fluctuations, which are exponentially rare in n (cyan area).

on this positivity, we have to understand the law of this smallest eigenvalue. More precisely, we will see that we must understand rare events in which the smallest eigenvalue can be very atypical, a regime called *large deviations* that we defined in Section 1.5.2 and that we illustrate in Fig. 7.2.

Applying Lemma 7.3 to eq. (7.5) with $t^2 = p/[(n-1)(p-1)]$ and $G = (t\sqrt{n})B$ yields:

$$\mathbb{E}\,\mathrm{Crit}^0_{n,p}(B) = 2\sqrt{\frac{2}{p}}(p-1)^{\frac{n}{2}}\,\mathbb{E}_{\mathrm{GOE}(n)}\left[e^{-n\frac{p-2}{4p}\lambda_0^2}\mathbb{1}\left(\lambda_0 \in \sqrt{\frac{p}{p-1}}B\right)\right]. \tag{7.6}$$

The large n limit

We are interested here in the *annealed* complexity $\Sigma^0_p(B) \equiv \lim_{n\to\infty}(1/n)\ln \mathbb{E}\,\mathrm{Crit}^0_{n,p}(B)$, as a first step towards understanding the landscape's topology. As we noticed above, this requires understanding the *large deviations* of the smallest eigenvalue of a GOE(n) matrix, cf. Fig. 7.2. This is precisely given by Theorem A.1 of [ABAČ13]:

Lemma 7.4 (*LDP of the smallest eigenvalue of a GOE matrix*)

Let λ_0 be the smallest eigenvalue of a GOE(n) matrix. The law of λ_0 satisfies a large deviation principle in the scale n, with rate function $I(x)$:

$$I(x) \equiv \begin{cases} +\infty & \text{if } x \geq -2, \\ \displaystyle\int_2^{-x} \mathrm{d}z \sqrt{\frac{z^2}{4} - 1} & \text{otherwise.} \end{cases}.$$

In Chapter 8 we will specifically focus on the large deviations of the smallest eigenvalue of random matrices, from an ensemble much larger than only the GOE. Using Lemma 7.4 alongside eq (7.6) and Varadhan's lemma 1.10 yields:

$$\Sigma^0_p(B) = \frac{1}{2}\ln(p-1) + \sup_{x \in \sqrt{\frac{p}{p-1}}B} \left[-\frac{p-2}{4p}x^2 - I(x)\right]. \tag{7.7}$$

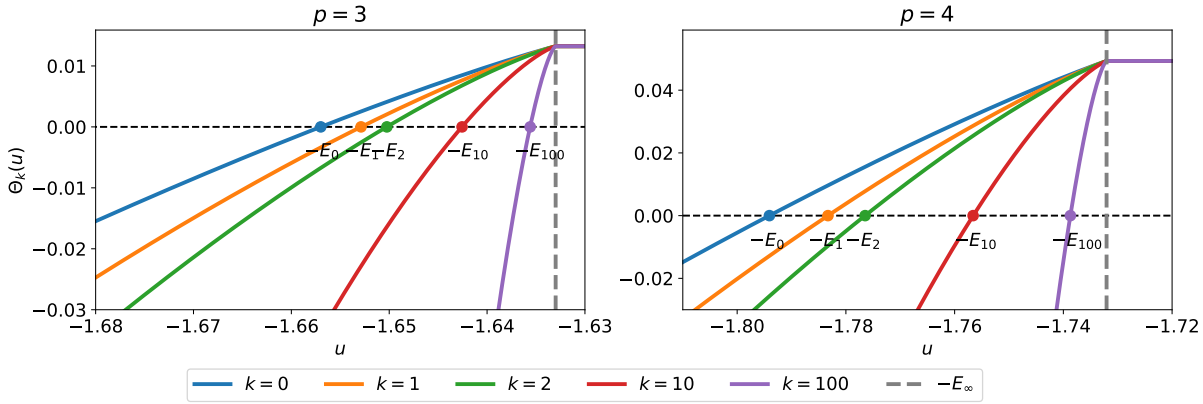


FIGURE 7.3: The functions Θ_k for $k \in 0, 1, 2, 10, 100$, for two values of p.

Description of the results

Considering $B = (-\infty, u)$ (for $u \in \mathbb{R}$) in eq. (7.7) amounts to count the local minima of the pure p-spin Hamiltonian of extensive energy smaller than nu. In this case, the supremum in eq. (7.7) can be analytically performed. Moreover, the calculation we sketched for local minima can be generalized for critical points of any fixed index $k \in \mathbb{N}$, and if we define:

$$\Theta_k(u) \equiv \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\, \mathrm{Crit}^k_{n,p}((-\infty, u)),$$

there exists analytic expressions for all $\Theta_k(u)$ functions, given in eq (2.16) of [ABAČ13]. We plot these functions for $p = 3, 4$ in Fig. 7.3. In particular, all these functions agree for $u \geq -E_\infty$, which is often referred to as the *threshold* energy

$$E_\infty \equiv 2\sqrt{\frac{p-1}{p}}. \tag{7.8}$$

We also denote $-E_k$ the value at which the function $\Theta_k(u)$ becomes positive: $\Theta_k(-E_k) = 0$, cf Fig. 7.3. We end this description by a series of remarks that one can make from the figure:

(*i*) The local minima always dominate the complexity for all energies below $-nE_\infty$. For $u \geq -E_\infty$, one can show that the complexity is dominated instead by critical points of diverging index.

(*ii*) The value $-E_0$ is the lowest energy of critical points, and corresponds to the ground state of the original function. It can be indeed shown (cf. [ABAČ13]) that the global minimum energy concentrates to a deterministic value as $n \to \infty$, and that this value is precisely $-E_0$!

(*iii*) One can perform similar calculations for the complexity of critical points whose indices diverge with n, see [ABA13] for the rigorous derivation.

To conclude our analysis, let us mention a few generalizations and extensions of the Kac-Rice calculation for p-spin models that we described:

- Beyond the annealed complexity, one can also use the Kac-Rice formula for higher-order moments, in order to compute e.g. the second moment of the number of critical points. For the pure spherical p-spin this is done in [Sub17a], which then shows by a second-moment method that the number of local minima concentrates. The *annealed* and *quenched* complexities of local minima are therefore identical for a pure spherical p-spin, which increases a lot the physical relevance of our previous discussion: we were describing the actual quenched complexity!

This allows to describe very detailed properties of the Gibbs measure at low temperatures, see e.g. [Sub17b, SZ17, BASZ20].

- Going further, one can even compute all the integer moments: this allows to perform heuristic replica calculations to obtain the quenched complexity, as performed for Gaussian models in [RBABC19].
- All the calculations we mentioned can be generalized to *mixed spherical p-spins* (cf. Model 1.4), which was done on [ABA13, RBABC19], or to bipartite spherical models (both pure and mixed) [McK21a].
- Our arguments relied quite strongly on the *isotropy* of the random field we are describing. Recent works in the physics and mathematics literature derive important extensions of the Kac-Rice approach to non-isotropic Gaussian model (e.g. the elastic manifold) [FLD20a, FLD20b, BABM21b].

Let us emphasize once again that the Kac-Rice method we presented transforms our original counting problem into a random matrix theory problem. Moreover, as we saw, counting critical points of specific index requires knowing the *large deviations* of the corresponding eigenvalues of the Hessian matrix. This last remark is important, as it will be one of the motivations behind Chapter 8.

7.2 Kac-Rice for inference models: main results

In the rest of this chapter, based on [MBAB20], we present an extension of the Kac-Rice method to compute the number of critical points of the empirical risk arising in generalized linear estimation problems, i.e. the GLMs that we already encountered several times in this thesis. We obtain a rigorous explicit variational formula for the *annealed complexity* (Theorems 7.5 and 7.6), that we then simplify and extend using a heuristic Kac-Rice replicated method, which originated in theoretical physics. In this way we find an explicit variational formula for the *quenched complexity* (Results 7.1 and 7.2), which allows to obtain the number of critical points for typical instances up to exponential accuracy.

Let us first mention a recent line of work that generalized the Kac-Rice calculations of Section 7.1 to a class of inference models known as *spiked matrix-tensor models* [SMKUZ19, SMBC$^+$19]. The authors leveraged the Kac-Rice formula and a precise random matrix analysis of the Hessian to gain precise insight into the landscape and the behavior of local optimization algorithms. Crucially, the loss function considered is still a *Gaussian random function* (although much more involved than the pure p-spin of eq. (7.3)), so that the random matrix problem resulting from the application of Kac-Rice is essentially the analysis of a (spiked) GOE matrix. We refer to [MZ20] for a review on this approach.

In this chapter we aim at going beyond this Gaussian setting, and we instead consider two classes of high-dimensional random functions.

- The first one is a kind of energy that arises in a simple model of neural networks (the perceptron, cf. [EVdB01]) and in mean-field glass models [FP16]:

$$L_1(\mathbf{x}) \equiv \frac{1}{m}\sum_{\mu=1}^{m} \phi(\boldsymbol{\xi}_\mu \cdot \mathbf{x}), \tag{7.9}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a "smooth" (the precise sense will be given later) activation function, $\mathbf{x} \in \mathbb{S}^{n-1}$ (the unit sphere in n dimensions) and $\boldsymbol{\xi}_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_n)$.

- The second class of functions we will consider is related to the loss functions of generalized linear models (GLMs), that we introduced in Section 1.1 and that we already investigated many times in this thesis. In the current setting, an observer has to infer a hidden vector $\mathbf{x}^\star \in \mathbb{S}^{n-1}$ from the observation of the m-dimensional output vector $\mathbf{Y} = \{\phi(\boldsymbol{\xi}_\mu \cdot \mathbf{x}^\star)\}_{\mu=1}^m$[5]. The data (or measurement) matrix $\boldsymbol{\xi}$ is taken random, with an i.i.d. standard Gaussian distribution, and we assume that the function ϕ and the data matrix $\boldsymbol{\xi}$ are given to the observer. This naturally leads to the mean square loss L_2:

$$L_2(\mathbf{x}) \equiv \frac{1}{2m} \sum_{\mu=1}^{m} \left[\phi(\boldsymbol{\xi}_\mu \cdot \mathbf{x}^\star) - \phi(\boldsymbol{\xi}_\mu \cdot \mathbf{x})\right]^2. \tag{7.10}$$

As we discussed in Section 7.1, we are interested in the statistics of the number of critical points of the functions of eqs. (7.9),(7.10). To introduce the notations we focus on L_2. For any open intervals $B \subseteq \mathbb{R}_+$ and $Q \subseteq (-1, 1)$, we consider the (random) number $\mathrm{Crit}_{n,L_2}(B, Q)$ of critical points of the function L_2 with loss value in B and overlap with the signal $q \equiv \mathbf{x} \cdot \mathbf{x}^\star$ in Q:

$$\mathrm{Crit}_{n,L_2}(B, Q) \equiv \sum_{\mathbf{x}:\, \mathrm{grad}\, L_2(\mathbf{x})=0} \mathbb{1}\{L_2(\mathbf{x}) \in B,\ \mathbf{x} \cdot \mathbf{x}^\star \in Q\}. \tag{7.11}$$

Here grad is the *Riemannian* gradient on $\mathbb{S}^{n-1}$. For L_1 we define the similar quantity $\mathrm{Crit}_{n,L_1}(B)$, dropping the notion of overlap.

Our main results consist in explicit formulas for the *annealed* and *quenched*[6] complexities for L_1 and L_2. The annealed formula is obtained rigorously using the Kac-Rice formula, whereas the quenched one is derived combining the Kac-Rice formula with the replica method, one of the classical tools of our statistical physics toolbox, cf Section 1.3.1. The quenched calculation leverages in particular important ideas of [RBABC19].

We consider the thermodynamic limit $n, m \to \infty$ with $m/n \to \alpha > 1$. The condition $\alpha > 1$ is essential, as can be seen e.g. in eq. (7.9): if $m < n$, for each realization of $\{\boldsymbol{\xi}_\mu\}$, the function L_1 has an infinite number of critical points in the set of vectors $\mathbf{x}$ orthogonal to all the $\{\boldsymbol{\xi}_\mu\}$, and counting the critical points in this case is meaningless (or one would have to quotient the space to lift the degeneracy).

Our results hold for many classical activation functions ϕ, such as e.g. the hyperbolic tangent, the arctangent, the sigmoid, or a smoothed and leaky version of the ReLU activation function[7].

For two probability measures μ, ν recall that we define the relative entropy (or Kullback-Leibler divergence) as $D_{\mathrm{KL}}(\mu|\nu) \equiv \int \ln(\mathrm{d}\mu/\mathrm{d}\nu)\mathrm{d}\mu$ if μ is absolutely continuous with respect to ν, and $+\infty$ otherwise. Finally, μ_G is a generic notation for the standard Gaussian measure on any $\mathbb{R}^k$.

The annealed complexity

We can now present our main results for the annealed complexity.

[5] In general GLMs, the output function is stochastic. Here, we restrict to deterministic outputs.
[6] Recall that we discussed the difference between annealed and quenched complexities in Section 7.1, cf eq. (7.1).
[7] The precise hypotheses on the activation function ϕ are precised in Section 7.3.

Theorem 7.5 (*Annealed complexity of L_1*)

Let $B \subseteq \mathbb{R}$ a non-empty open interval and denote $\mathcal{M}_\phi(B)$ the set of probability measures ν on $\mathbb{R}$ such that $\int \nu(\mathrm{d}t)\phi(t) \in B$. We define:

- $\mathcal{E}_\phi(\nu) \equiv \ln\big[\int \nu(\mathrm{d}x)\phi'(x)^2\big]$,
- $t_\phi(\nu) \equiv \int \nu(\mathrm{d}x)x\phi'(x)$,
- Let $\mathbf{z} \in \mathbb{R}^{n\times m}$ an i.i.d. standard Gaussian matrix, and $\mathbf{y} \in \mathbb{R}^m$ a vector with components taken i.i.d. from a probability measure ν. Let $\mathbf{D}^{(\nu)}$ the diagonal matrix of size m with elements $D^{(\nu)}_\mu = \phi''(y_\mu)$. We define $\mu_{\alpha,\phi}[\nu]$ as the LSD of $\mathbf{z}\mathbf{D}^{(\nu)}\mathbf{z}^\intercal/m$.
- $\kappa_{\alpha,\phi}(\nu, C) \equiv \int \mu_{\alpha,\phi}[\nu](\mathrm{d}x)\ln|x - C|$,

Under a technical assumption (see the remark below), one has[8]:

$$\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\ \mathrm{Crit}_{n,L_1}(B) = \frac{1+\ln\alpha}{2} + \sup_{\nu\in\mathcal{M}_\phi(B)}\left\{-\frac{1}{2}\mathcal{E}_\phi(\nu) + \kappa_{\alpha,\phi}(\nu, t_\phi(\nu)) - \alpha D_{\mathrm{KL}}(\nu|\mu_G)\right\}.$$

A note on free probability – Interestingly, the measure $\mu_{\alpha,\phi}[\nu]$ can be interpreted as the free multiplicative convolution of the Marchenko-Pastur law (at ratio α) and the asymptotic spectral distribution of the matrix $\mathbf{D}^{(\nu)}$, cf. e.g. [Voi87, AGZ10][9]. We describe in Section 7.4 how to explicitly compute the density of $\mu_{\alpha,\phi}[\nu]$, or its linear spectral statistics (as e.g. $\kappa_{\alpha,\phi}(\nu, C)$), via the computation of its Stieltjes transform.

We turn to our second (quite heavy) annealed result:

[8] A fully rigorous statement would imply a lower and an upper bound given by a supremum over the adherence and the interior of $\mathcal{M}_\phi(B)$. For reasons of lightness and clarity of the presentation we write it in the simpler presented form.

[9] Free multiplication is usually defined for positively-supported measures, however one can generalize it here by explicitly separating the positive and negative parts of ϕ'' (we can show freeness of the resulting two random matrices).

Theorem 7.6 (*The annealed complexity of L_2*)

Let $B \subseteq \mathbb{R}_+$ and $Q \subseteq (-1,1)$ two non-empty open intervals. For $q \in (-1,1)$ we denote $\mathcal{M}_\phi(B,q)$ the set of probability measures ν on $\mathbb{R}^2$ such that

$$\begin{cases} \int \nu(\mathrm{d}x,\mathrm{d}y)\, y\, \phi'(x)[\phi(qx+\sqrt{1-q^2}y)-\phi(x)] = 0, \\ \int \nu(\mathrm{d}x,\mathrm{d}y)[\phi(qx+\sqrt{1-q^2}y)-\phi(x)]^2 \in B. \end{cases}$$

Given the following definitions:

- $\mathcal{E}_\phi(q,\nu) \equiv \ln \int \nu(\mathrm{d}x,\mathrm{d}y)\, \phi'(x)^2 \big[\phi(qx+\sqrt{1-q^2}y)-\phi(x)\big]^2$,
- $t_\phi(q,\nu) \equiv \int \nu(\mathrm{d}x,\mathrm{d}y)\, x\, \phi'(x)\big[\phi(x)-\phi(qx+\sqrt{1-q^2}y)\big]$,
- $f_q(x,y) \equiv \phi'(x)^2 - \phi''(x)\big[\phi(qx+\sqrt{1-q^2}y)-\phi(x)\big]$,
- Let $\mathbf{z} \in \mathbb{R}^{n\times m}$ an i.i.d. standard Gaussian matrix, and $\mathbf{Y} \in \mathbb{R}^{m\times 2}$ with components taken i.i.d. from $\nu \in \mathcal{M}_1^+(\mathbb{R}^2)$. Let $\mathbf{D}^{(\nu,q)}$ the diagonal matrix of size m with elements $D_\mu^{(\nu,q)} = f_q(Y_\mu)$. We define $\mu_{\alpha,\phi}[q,\nu]$ as the LSD of $\mathbf{z}\mathbf{D}^{(\nu,q)}\mathbf{z}^\intercal/m$.
- $\kappa_{\alpha,\phi}(q,\nu) \equiv \int \mu_{\alpha,\phi}[q,\nu](\mathrm{d}x) \ln|x - t_\phi(q,\nu)|$.

Then one has[10]:

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\,\mathrm{Crit}_{n,L_2}(B,Q) = \frac{1+\ln\alpha}{2} + \sup_{q\in Q} \sup_{\nu\in\mathcal{M}_\phi(B,q)} \Big[\frac{1}{2}\ln(1-q^2) - \frac{1}{2}\mathcal{E}_\phi(q,\nu) + \kappa_{\alpha,\phi}(q,\nu) - \alpha D_{\mathrm{KL}}(\nu|\mu_G)\Big].$$

The proof of Theorem 7.5 will be the main subject of Section 7.3.

Technical assumption – Our proof relies on an assumption given in eq. (E.16). Under Definition 7.1 on the behavior of $\phi(x)$, we are currently working on proving this assumption by leveraging the recent work of close collaborators [BABM21a]. However, note that eq. (E.16) can also be considered as an hypothesis on $\phi(x)$, i.e. that there is no accumulation of eigenvalues around 0 in the spectrum of the Hessian of $L_1(\mathbf{x})$. This assumption is quite natural, and we found it to be satisfied for all relevant functions $\phi(x)$ we investigated, so that it does not limit the reach of our results.

The proof of Theorem 7.6 is a straightforward generalization, and we will sketch it briefly in Appendix E.2.5.

The variational problems in Theorems 7.5 and 7.6 are challenging, as they imply an optimization on a set of measures, and they involve transforms of this measure that are very hard to access numerically. In Section 7.4 we present a drastic simplification: a heuristic calculation that allows one to reduce the supremum over the probability measure ν to a much simpler optimization over a relatively small number of parameters.

The quenched complexity

As we have already stressed, the annealed complexity, although interesting in itself, is generically not representative of the landscape corresponding to a given typical instance of the empirical risk. In order to obtain the value of the quenched complexity we use the *replicated Kac-Rice*

[10] We use the same technical assumption

method, which is an extension to non-Gaussian functions of the one developed in [RBABC19]. As we know, although the replica method is non-rigorous it has been proven to give exact results for both spin glasses and inference problems [Tal06, BKM+19] (and proving the replica conjectures was one of the main aims of Part II of this thesis.). We have obtained an explicit formula[11] for the *quenched complexity* of L_1 and L_2 at fixed values of the empirical risk, and overlap with the solution (in the L_2 case).

For L_1, using the notations of Theorem 7.5 we have:

Result 7.1 (*Quenched complexity of L_1*)

Let $B \subseteq \mathbb{R}$ an open interval. Then:

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E} \ln \mathrm{Crit}_{n,L_1}(B) = \frac{\ln\alpha - \alpha \ln 2\pi}{2} + \sup_{\substack{\nu \in \mathcal{M}_\phi(B) \\ q\in(0,1)}} \mathrm{extr} \Bigg\{ \kappa_{\alpha,\phi}(\nu, C) + \frac{1-\alpha}{2}\ln(1-q)$$

$$+ \frac{1-\alpha q}{2(1-q)} - \int \nu(\mathrm{d}\lambda) g(\lambda) - \frac{A\hat{A} - a\hat{a}}{2} + C(q\hat{c} - \hat{C}) - \frac{1}{2}\ln[A-a] - \frac{a}{2(A-a)} + \alpha \int_{\mathbb{R}^4} \mathcal{D}\boldsymbol{\xi} \ln I(\boldsymbol{\xi}) \Bigg\}.$$

Here $\boldsymbol{\xi} \equiv (\xi_q, \xi_a, \xi_c, \xi_c')$ and $\mathcal{D}\boldsymbol{\xi}$ is the standard Gaussian probability measure on $\mathbb{R}^4$. extr denotes extremization with respect to all variables $(A, \hat{A}, a, \hat{a}, C, \hat{C}, \hat{c}, \{g(\lambda)\})$. We defined

$$I(\boldsymbol{\xi}) \equiv \int_{\mathbb{R}} \mathrm{d}\lambda\, e^{-\frac{\lambda^2}{2(1-q)} + \frac{g(\lambda)}{\alpha} + \frac{\hat{A}-\hat{a}}{2\alpha}\phi'(\lambda)^2 + \frac{\hat{C}-\hat{c}}{\alpha}\phi'(\lambda)\lambda + \frac{\sqrt{q}}{1-q}\xi_q\lambda + \sqrt{\frac{\hat{a}}{\alpha}}\xi_a\phi'(\lambda) + \sqrt{\frac{\hat{c}}{2\alpha}}[\phi'(\lambda)(\xi_c + i\xi_c') + \lambda(\xi_c - i\xi_c')]}.$$

On the extremization –In this formula, the notation "extr" denotes that one should set the partial derivatives with respect to the involved variables to zero. This notation arises from the replica calculation, which mixes saddle-point computations with Lagrange multipliers associated to certain constraints, and the precise meaning of this extremization (as a supremum or infimum) would have to be clarified by a more rigorous method. On a numerical point of view, one would have to solve the associated saddle-point equations, so that this precise meaning is not crucial for applications.

We can state a very similar result for L_2:

[11] We used a replica symmetric (RS) structure, which is correct in many cases, and a very good approximation in others were replica symmetry has to be broken. See Section 1.3.1 for more discussion on the RS ansatz.

Result 7.2 (*Quenched complexity of L_2*)

Let $B \subseteq \mathbb{R}, Q \subseteq (-1,1)$ two open intervals and define:

- For $m \in (-1,1)$, $\mathcal{M}_\phi(B,m)$ is the set of $\nu \in \mathcal{M}_1^+(\mathbb{R}^2)$ such that:

$$\begin{cases} \frac{1}{2}\int \nu(\mathrm{d}\lambda^0, \mathrm{d}\lambda)[\phi(\lambda)-\phi(\lambda^0)]^2 \in B, \\ \int \nu(\mathrm{d}\lambda^0, \mathrm{d}\lambda)\phi'(\lambda)[\phi(\lambda)-\phi(\lambda^0)](\lambda^0 - m\lambda) = 0. \end{cases}$$

- Let $f(x,y) \equiv \phi''(y)[\phi(y)-\phi(x)] + \phi'(y)^2$. Let $\mathbf{z} \in \mathbb{R}^{n\times m}$ an i.i.d. standard Gaussian matrix, and $\mathbf{Y} \in \mathbb{R}^{m\times 2}$ with components taken i.i.d. from $\nu \in \mathcal{M}_1^+(\mathbb{R}^2)$. Let $\mathbf{D}^{(\nu)}$ the diagonal matrix of size m with elements $D_\mu^{(\nu)} = f(Y_\mu)$. We define $\mu_{\alpha,\phi}[\nu]$ as the asymptotic spectral measure of $\mathbf{z}\mathbf{D}^{(\nu)}\mathbf{z}^\intercal/m$.

- $\chi_{\alpha,\phi}(\nu, C) \equiv \int \mu_{\alpha,\phi}[\nu](\mathrm{d}x) \ln|x-C|$.

One has:

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\ln \mathrm{Crit}_{n,L_2}(B,Q) = \sup_{\substack{m\in Q \\ q\in(0,1)}} \sup_{\nu\in\mathcal{M}_\phi(B,m)} \mathrm{extr}\Big\{ & \frac{\ln\alpha - \alpha\ln 2\pi}{2} + \chi_{\alpha,\phi}(\nu, C) \\ & + \frac{1-\alpha q - m^2}{2(1-q)} + \frac{1-\alpha}{2}\ln(1-q) - \frac{1}{2}\ln(A-a) - \frac{a}{2(A-a)} - \frac{A\hat{A}}{2} + \frac{a\hat{a}}{2} \\ & - C_0\hat{C}_0 - C\hat{C} + c\hat{c} - \int \nu(\mathrm{d}\lambda^0, \mathrm{d}\lambda) g(\lambda^0,\lambda) + \alpha \int_{\mathbb{R}^4\times\mathbb{R}} \mathcal{D}\boldsymbol{\xi}\, \mathcal{D}\lambda^0 \ln I(\lambda^0, \boldsymbol{\xi}) \Big\}. \end{aligned}$$

The extremum is made over all variables $(A, a, C_0, C, c, \hat{A}, \hat{a}, \hat{C}, \hat{c}, \hat{C}_0, \{g(\lambda^0,\lambda)\})$. $\mathcal{D}$ is the standard Gaussian measure, and the variables C_0, c, C are related by the additional constraint

$$-m(1-q)C_0 - (q-m^2)C + (1-m^2)c = 0.$$

$I(\lambda^0, \boldsymbol{\xi})$ is defined as (with $\boldsymbol{\xi} \equiv (\xi_q, \xi_a, \xi_c, \xi_c')$):

$$\begin{aligned} I(\lambda^0,\boldsymbol{\xi}) \equiv \int_{\mathbb{R}} \mathrm{d}\lambda\, e^{\frac{m}{1-q}\lambda^0\lambda - \frac{\lambda^2}{2(1-q)} + \frac{\sqrt{q-m^2}}{1-q}\xi_q\lambda + \frac{g(\lambda^0,\lambda)}{\alpha} + \frac{\hat{C}_0}{\alpha}\phi'(\lambda)[\phi(\lambda)-\phi(\lambda^0)]\lambda^0 + \frac{\hat{C}-\hat{c}}{\alpha}\lambda\phi'(\lambda)[\phi(\lambda)-\phi(\lambda^0)]} \\ e^{\frac{\hat{A}-\hat{a}}{2\alpha}\phi'(\lambda)^2[\phi(\lambda)-\phi(\lambda^0)]^2 + \sqrt{\frac{\hat{a}}{\alpha}}\xi_a\phi'(\lambda)[\phi(\lambda)-\phi(\lambda^0)] + \sqrt{\frac{\hat{c}}{2\alpha}}\left[\phi'(\lambda)[\phi(\lambda)-\phi(\lambda^0)](\xi_c+i\xi_c') + \lambda(\xi_c - i\xi_c')\right]}. \end{aligned}$$

The derivation of Result 7.1 is given in Section 7.5. Similarly to the annealed case, Result 7.2 can be derived by a simple generalization of this computation, see Appendix E.2.5.

Tackling the variational problems

- As a sanity check of our results, the reader can analytically check by explicit solution that for a linear activation function, the annealed complexity of L_1 is null in Theorem 7.5. It is again a tedious but straightforward computation to check that the annealed complexity of L_1 with a quadratic activation $\phi(x) = x^2$ is also null, as the number of critical points in this case is linear with n. In both these cases, the measure $\mu_{\alpha,\phi}[\nu]$ is indeed independent of ν, which simplifies the problem drastically. Note that for L_2 however, even the case of a linear activation is non trivial (because of the spherical constraint), as shown in the recent analysis of [FT20].

- However, for more generic activation functions, solving the variational problem of the previous annealed and quenched calculations is a very involved task. We explain in Section 7.4 a route

to make tractable theses variational problems, by reducing them to optimization problems over a relatively small number of parameters.

- Apart from the simple cases described above, it is not clear to deduce from our results if the quenched and annealed complexities are equal for given values of the loss function, or for a given activation function. Such an equality is equivalent to the concentration of the number of critical points, and its occurrence (or lack thereof) has deep physical consequences on our understanding of these landscapes. Note that while this happens notably in the pure p-spin, as shown by a second moment analysis in [Sub17a], it remains a very peculiar situation, not satisfied e.g. by generic spherical p-spin "mixtures", and in the context of generalized linear models it would have to be investigated through an extensive numerical analysis of our results, which is not presented in this thesis.

7.3 Proof of the annealed complexity

In this section we prove Theorem 7.5. Our technique leverages the Kac-Rice formula and Sanov's theorem 1.9 on the large deviations of the empirical measure of i.i.d. variables. First we precise our hypotheses on ϕ, that we will take in the following set of "well-behaved" activation functions:

Definition 7.1 (*Well-behaved activation function*)

$\phi : \mathbb{R} \to \mathbb{R}$ is "well-behaved" if it is of class $\mathcal{C}^3$ and if, for $y \sim \mathcal{N}(0,1)$, the random variable $a = \phi'(y)$ admits a continuous probability density in a neighborhood of $a = 0$.

7.3.1 Applying the Kac-Rice formula

The first step is to apply the Kac-Rice formula to the random function L_1:

Lemma 7.7 (*Kac-Rice formula for L_1*)

For any $\mathbf{x} \in \mathbb{S}^{n-1}$, denote $\operatorname{grad} L_1(\mathbf{x})$ and $\operatorname{Hess} L_1(\mathbf{x})$ the (Riemannian) gradient and Hessian of L_1 at the point $\mathbf{x}$. Then $\operatorname{grad} L_1(\mathbf{x})$ has a well defined density (on the tangent space $T_{\mathbf{x}}\mathbb{S}^{n-1} \simeq \mathbb{R}^{n-1}$) in a neighborhood of zero, that we denote $\varphi_{\operatorname{grad} L_1(\mathbf{x})}$. Denote $\mu_{\mathbb{S}}$ the usual surface measure on $\mathbb{S}^{n-1}$. One has:

$$\mathbb{E}\operatorname{Crit}_{n,L_1}(B) = \int_{\mathbb{S}^{n-1}} \varphi_{\operatorname{grad} L_1(\mathbf{x})}(0)\mathbb{E}\big[\mathbb{1}\{L_1(\mathbf{x}) \in B\}|\det \operatorname{Hess} L_1(\mathbf{x})|\big|\operatorname{grad} L_1(\mathbf{x}) = 0\big]\mu_{\mathbb{S}}(\mathrm{d}\mathbf{x}).$$

This lemma is a direct application of Proposition 7.2 and uses necessary conditions for a random function of the type of L_1 to be a.s. Morse that are stated in [AW09b]. The details of the proof are given in Appendix E.2.1.

7.3.2 The complexity at finite n

For $\mathbf{y} \in \mathbb{R}^m$, let $\mathbf{\Lambda}(\mathbf{y}) \in \mathcal{S}_m$ be

$$\mathbf{\Lambda}(\mathbf{y}) \equiv \Big(\mathrm{I}_m - \frac{\phi'(\mathbf{y})\phi'(\mathbf{y})^\intercal}{\|\phi'(\mathbf{y})\|^2}\Big)\mathbf{D}(\mathbf{y})\Big(\mathrm{I}_m - \frac{\phi'(\mathbf{y})\phi'(\mathbf{y})^\intercal}{\|\phi'(\mathbf{y})\|^2}\Big), \tag{7.12}$$

in which we denote $\phi'(\mathbf{y}) \equiv (\phi'(y_\mu))_{\mu=1}^m$, and $\mathbf{D}(\mathbf{y}) \in \mathbb{R}^{m\times m}$ the diagonal matrix with elements $\mathbf{D}(\mathbf{y})_\mu = D(y_\mu) = n\phi''(y_\mu)/m$. We aim at proving the following lemma:

Lemma 7.8 (*Complexity at finite n*)

$$\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B) = \mathcal{C}_n\, e^{n\frac{1+\ln\alpha}{2}}\mathbb{E}_{\mathbf{y}}\left[\mathbb{1}_{\frac{1}{m}\sum_\mu \phi(y_\mu)\in B}\; e^{-\frac{n-1}{2}\ln\left(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2\right)}\mathbb{E}_{\mathbf{z}}\left[\left|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})\right|\right]\right],$$

in which $\mathcal{C}_n$ is *exponentially trivial*, i.e. $(1/n)\ln\mathcal{C}_n = \mathcal{O}_n(1)$. The variable $\mathbf{y}\in\mathbb{R}^m$ follows $\mathcal{N}(0,\mathrm{I}_m)$, and $\mathbf{z}\in\mathbb{R}^{(n-1)\times m}$ has i.i.d. standard Gaussian matrix elements, independent of $\mathbf{y}$. $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$ is a square matrix of size $(n-1)$ with the following distribution :

$$\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y}) \overset{\mathrm{d}}{=} \frac{1}{n}\mathbf{z}\mathbf{\Lambda}(\mathbf{y})\mathbf{z}^\intercal - \Big\{\frac{1}{m}\sum_{\mu=1}^m y_\mu\phi'(y_\mu)\Big\}\mathrm{I}_{n-1}. \tag{7.13}$$

The rest of Section 7.3.2 is devoted to the proof of Lemma 7.8. First, the following proposition specifies the joint distribution of $(L_1(\mathbf{x}), \operatorname{grad} L_1(\mathbf{x}), \operatorname{Hess} L_1(\mathbf{x}))$, which will be useful to apply Lemma 7.7.

Proposition 7.9 (*Distribution of the gradient and Hessian*)

Let $\mathbf{x}\in\mathbb{S}^{n-1}$. Then $(L_1(\mathbf{x}), \operatorname{grad} L_1(\mathbf{x}), \operatorname{Hess} L_1(\mathbf{x}))$ follows the following joint distribution:

$$\begin{cases} L_1(\mathbf{x}) \overset{\mathrm{d}}{=} \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m \phi(y_\mu), & \text{(7.14a)}\\[2ex] \operatorname{grad} L_1(\mathbf{x}) \overset{\mathrm{d}}{=} \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m \phi'(y_\mu)\mathbf{z}_\mu, & \text{(7.14b)}\\[2ex] \operatorname{Hess} L_1(\mathbf{x}) \overset{\mathrm{d}}{=} \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m \phi''(y_\mu)\mathbf{z}_\mu\mathbf{z}_\mu^\intercal - \Big\{\dfrac{1}{m}\displaystyle\sum_{\mu=1}^m y_\mu\phi'(y_\mu)\Big\}\mathrm{I}_{n-1}, & \text{(7.14c)} \end{cases}$$

in which $\mathbf{y} = (y_\mu)_{\mu=1}^m \sim \mathcal{N}(0,\mathrm{I}_m)$, $(\mathbf{z}_\mu)_{\mu=1}^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\mathrm{I}_{n-1})$, and all $\{y_\mu, \mathbf{z}_\nu\}$ are independent. We identified in these equations the tangent space $T_{\mathbf{x}}\mathbb{S}^{n-1}$ with $\mathbb{R}^{n-1}$.

Proof of Proposition 7.9 – Denote $P_{\mathbf{x}}^\perp$ the orthogonal projection on $\{\mathbf{x}\}^\perp$. For a smooth function $f:\mathbb{S}^{n-1}\to\mathbb{R}$, ∇f and $\nabla^2 f$ denote its *Euclidean* gradient and Hessian. The Riemannian structure on $\mathbb{S}^{n-1}$ induces the gradient and Hessian of f as $\operatorname{grad} f(\mathbf{x}) = P_{\mathbf{x}}^\perp\nabla f$ and $\operatorname{Hess} f(\mathbf{x}) = P_{\mathbf{x}}^\perp\nabla^2 f P_{\mathbf{x}}^\perp - (\mathbf{x}\cdot\nabla f(\mathbf{x}))P_{\mathbf{x}}^\perp$. Applying these formulas yields:

$$\begin{cases} \operatorname{grad} L_1(\mathbf{x}) = \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m (P_{\mathbf{x}}^\perp\boldsymbol{\xi}_\mu)\phi'(\boldsymbol{\xi}_\mu\cdot\mathbf{x}), & \text{(7.15a)}\\[2ex] \operatorname{Hess} L_1(\mathbf{x}) = \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m \phi''(\boldsymbol{\xi}_\mu\cdot\mathbf{x})(P_{\mathbf{x}}^\perp\boldsymbol{\xi}_\mu)(P_{\mathbf{x}}^\perp\boldsymbol{\xi}_\mu)^\intercal - \Big\{\dfrac{1}{m}\displaystyle\sum_{\mu=1}^m (\boldsymbol{\xi}_\mu\cdot\mathbf{x})\phi'(\boldsymbol{\xi}_\mu\cdot\mathbf{x})\Big\}P_{\mathbf{x}}^\perp. & \text{(7.15b)} \end{cases}$$

Letting $y_\mu \equiv \boldsymbol{\xi}_\mu\cdot\mathbf{x}$ and $\mathbf{z}_\mu \equiv P_{\mathbf{x}}^\perp\boldsymbol{\xi}_\mu$ (identified to an element of $\mathbb{R}^{n-1}$) yields the result. □

The joint distribution of eq. (7.14) does not depend on $\mathbf{x}$, thus we can chose $\mathbf{x}$ to be the North pole $\mathbf{x} = \mathbf{e}_n = (\delta_{i,n})_{i=1}^n$. With $\omega_n \equiv 2\pi^{n/2}/\Gamma(n/2)$ the volume of $\mathbb{S}^{n-1}$, we obtain from Lemma 7.7:

$$\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B) = \omega_n\varphi_{\operatorname{grad} L_1(\mathbf{e}_n)}(0)\mathbb{E}\big[|\det\operatorname{Hess} L_1(\mathbf{e}_n)|\mathbb{1}_{L_1(\mathbf{e}_n)\in B}\big|\operatorname{grad} L_1(\mathbf{e}_n) = 0\big]. \tag{7.16}$$

Removing the $\mathbf{e}_n$ indication and conditioning on the distribution of $\mathbf{y}$, we reach:

$$\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B) = \omega_n\mathbb{E}_{\mathbf{y}}\Big\{\mathbb{1}_{\frac{1}{m}\sum_\mu\phi(y_\mu)\in B}\,\varphi_{\operatorname{grad}L_1|\mathbf{y}}(0)\,\mathbb{E}_{\mathbf{z}}\big[|\det\operatorname{Hess} L_1|\big|\operatorname{grad} L_1 = 0,\mathbf{y}\big]\Big\}.$$

Once conditioned on $\mathbf{y}$, eq. (7.14b) describes a Gaussian density so we can directly compute:

$$\begin{aligned}\omega_n \varphi_{\mathrm{grad} L_1|\mathbf{y}}(0) &= \frac{2\pi^{n/2}}{\Gamma(n/2)} \exp\left[-\frac{n-1}{2} \ln\left(\frac{2\pi}{m^2} \sum_{\mu=1}^{m} \phi'(y_\mu)^2 \right) \right], \\ &= \mathcal{C}_n \exp\left\{ \frac{n}{2} + \frac{n}{2} \ln \frac{m}{n} - \frac{n-1}{2} \ln\left(\frac{1}{m} \sum_{\mu=1}^{m} \phi'(y_\mu)^2 \right) \right\}, \end{aligned} \tag{7.17}$$

in which $\ln \mathcal{C}_n = \mathcal{O}_n(n)$ (using Stirling's formula). The conditioning of $\mathrm{Hess}\, L_1$ by $\mathrm{grad}\, L_1 = 0$ at fixed $\mathbf{y}$ reduces to a linear conditioning on $\mathbf{z}$. One thus obtains by classical Gaussian conditioning:

$$\mathbb{E}_{\mathbf{z}}\big[|\det \mathrm{Hess} L_1 | \big| \mathrm{grad}\, L_1 = 0, \mathbf{y} \big] = \mathbb{E}_{\mathbf{z}}\big[|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| \big], \tag{7.18}$$

in which $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$ is defined by eq. (7.13). This ends the proof of Lemma 7.8.

7.3.3 Concentration and large deviations

This section is devoted to the end of the proof of Theorem 7.5. We denote $\nu_{\mathbf{y}}^m \equiv m^{-1} \sum_{\mu=1}^m \delta_{y_\mu}$ the empirical distribution of $\mathbf{y}$, and take the notations of Theorem 7.5 and Lemma 7.8. We first state an important lemma on the concentration of $\mathbb{E}_{\mathbf{z}}\big[|\det \mathbf{H}_n^{\Lambda}(\mathbf{y})| \big]$[12]:

Lemma 7.10 (*Concentration of the log-determinant*)

There exists $\eta > 0$ such that for all $t > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}\Big[\Big| \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}}\big[|\det \mathbf{H}_n^{\Lambda}(\mathbf{y})| \big] - \kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m)) \Big| \geq t \Big] = -\infty.$$

The proof of Lemma 7.10 is detailed in Appendix E.2.2, and is the section of our proof that requires to assume eq. (E.16). On a general note, we expect this result to be valid in the whole range $\eta \in (0,1)$, as the large deviations of the spectral distribution of random matrices is typically on the n^2 scale [BAG97, HP98]. Note that very similar results on the concentration of determinants of very generic classes of random matrices have recently been analyzed in [BABM21a] (see as well the companion papers [BABM21b, McK21a]). The following moment condition, proven in Appendix E.2.3, will be important to apply Varadhan's lemma 1.10:

Lemma 7.11

For every $\gamma \in (1, \alpha)$ we have:

$$\begin{cases} \limsup_{n\to\infty} \dfrac{1}{n} \ln \mathbb{E}_{\mathbf{y}} \left\{ e^{\gamma n \left[-\frac{1}{2} \ln \left(\frac{1}{m} \sum_\mu \phi'(y_\mu)^2 \right) + \kappa_{\alpha,\phi}\left(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m) \right) \right]} \right\} < +\infty, & \text{(7.19a)} \\ \limsup_{n\to\infty} \dfrac{1}{n} \ln \mathbb{E}_{\mathbf{y}} \left\{ e^{\gamma n \left[-\frac{1}{2} \ln \left(\frac{1}{m} \sum_\mu \phi'(y_\mu)^2 \right) + \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}} \left[|\det \mathbf{H}_n^{\Lambda}(\mathbf{y})| \right] \right]} \right\} < +\infty. & \text{(7.19b)} \end{cases}$$

We then conclude from Lemmas 7.8, 7.10 and 7.11, using Sanov's principle (Theorem 1.9) and Varadhan's lemma 1.10. We finally reach the statement of Theorem 7.5:

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\, \mathrm{Crit}_{n,L_1}(B) = \sup_{\nu \in \mathcal{M}_\phi(B)} \Big[\frac{1 + \ln \alpha}{2} - \frac{\mathcal{E}_\phi(\nu)}{2} + \kappa_{\alpha,\phi}(\nu, t_\phi(\nu)) - \alpha D_{\mathrm{KL}}(\nu | \mu_G) \Big]. \tag{7.20}$$

[12] In the proofs of this section we assume that $x\phi'(x)$ and $\phi''(x)$ are bounded. As one can always smoothly truncate the largest values of ϕ without affecting the complexity, this does not remove any generality to our results.

The detailed proof of eq. (7.20) from the lemmas above is given in Appendix E.2.4.

7.4 Towards a numerical solution?

7.4.1 The logarithmic potential of $\mu_{\alpha,\phi}[\nu]$

Let $\nu \in \mathcal{M}_1^+(\mathbb{R})$. As we have shown in the introduction to this thesis, cf. Theorem 1.7, the Stieltjes transform $g(z) \equiv \int \mu(\mathrm{d}t)(t-z)^{-1}$ of $\mu_{\alpha,\phi}[\nu]$ is given by the unique solution in $\mathbb{C}_+$ to the implicit *Marchenko-Pastur* equation:

$$\forall z \in \mathbb{C}_+, \quad g(z) = -\Big[z - \alpha \int \frac{\phi''(t)}{\alpha + \phi''(t) g(z)} \nu(\mathrm{d}t)\Big]^{-1}. \tag{7.21}$$

For any $\mu \in \mathcal{M}_1^+(\mathbb{R})$ and $t \in \mathbb{R}$ we define the *logarithmic potential* as $U[\mu](t) \equiv \int \mu(\mathrm{d}x) \ln|x-t|$. It is well defined with values in $\mathbb{R} \cup \{\pm\infty\}$, and the reader can refer to [Far14] for a review on this subject. To numerically evaluate Theorem 7.5, we have to compute $U[\mu](t)$ for $\mu = \mu_{\alpha,\phi}[\nu]$ and an arbitrary $t \in \mathbb{R}$. For clarity, we will write $\mu = \mu_{\alpha,\phi}[\nu]$ for the remainder of this section. Let us define $G(z) \equiv \int \mu(\mathrm{d}x) \ln(z-x)$ for any $z \in \mathbb{C}_+$. One sees directly that $G(z)$ is holomorphic on $\mathbb{C}_+$. Moreover, from Chapter II of [Far14], we know that $U[\mu](t) = \lim_{\epsilon \downarrow 0} \mathrm{Re}\, G(t + i\epsilon)$.

It is then clear that a way to compute the logarithmic potential is to evaluate $G(z)$ for $z \in \mathbb{C}_+$. Define, for $z, g \in \mathbb{C}_+$[13]:

$$F(z,g) \equiv -\ln(-g) - zg + \alpha \int \nu(\mathrm{d}\lambda) \ln(\alpha + \phi''(\lambda) g) - 1 - \alpha \ln \alpha. \tag{7.22}$$

At any fixed z, $F(z,g)$ is an holomorphic function of g on $\mathbb{C}_+$. Its Wirtinger derivative is:

$$\frac{\partial F}{\partial g}(z,g) = -\frac{1}{g} - z + \alpha \int \nu(\mathrm{d}\lambda) \frac{\phi''(\lambda)}{\alpha + \phi''(\lambda) g}.$$

Thus $g(z)$ (the Stieltjes transform of μ, cf. eq. (7.21)) is the *only* $g \in \mathbb{C}_+$ such that $\partial_g F(z,g) = 0$. Moreover, by definition $g(z)$ is an holomorphic function on $\mathbb{C}_+$, with values in $\mathbb{C}_+$. We can thus apply the usual composition of derivatives and obtain:

$$\frac{\mathrm{d}F}{\mathrm{d}z}(z, g(z)) = -g(z).$$

Furthermore by definition $\mathrm{d}G/\mathrm{d}z = -g(z)$. Computing the remaining constant by investigating the limit $\mathrm{Re}[z] \to \infty$, the reader can easily check that $G(z) = F(z, g(z))$ for every $z \in \mathbb{C}_+$. We thus have the crucial relation:

$$\forall t \in \mathbb{R}, \quad U[\mu_{\alpha,\phi}[\nu]](t) = \lim_{\epsilon \downarrow 0} \mathrm{Re}\, F[t + i\epsilon, g(t + i\epsilon)]. \tag{7.23}$$

This formula allows for an efficient numerical derivation of the logarithmic potential of $\mu_{\alpha,\phi}[\nu]$, for any value of t (including possibly inside the bulk of $\mu_{\alpha,\phi}[\nu]$), as we will detail below.

7.4.2 Heuristic derivation of simplified fixed point equations

We present here an heuristic derivation of scalar fixed point equations for the numerical resolution of Theorem 7.5. This technique can be easily extended to Theorem 7.6 as well as the quenched calculations presented afterwards, and we restrict to this simpler case for the sake of the presentation.

[13] $g \in \mathbb{C}_+$, and (since $\alpha > 1$) $\alpha + \phi''(\lambda) g \in \mathbb{C} \backslash (-\infty, 1]$, thus we can use the principal determination of the logarithm.

Expressing $\kappa_{\alpha,\phi}(\nu,t)$

From eq. (7.23) we know that for every $t\in\mathbb{R}$:

$$\kappa_{\alpha,\phi}(\nu,t)=\lim_{\epsilon\downarrow 0}\mathrm{Re}\,\big\{-\ln(-g(t+i\epsilon))-(t+i\epsilon)g(t+i\epsilon)+\alpha\int\nu(\mathrm{d}\lambda)\ln[\alpha+\phi''(\lambda)g(t+i\epsilon)]\\ -1-\alpha\ln\alpha\big\}.$$

Recall that $g(t+i\epsilon)$ is the only solution in $\mathbb{C}_+$ to the partial derivative of the previous equation:

$$-\frac{1}{g}-(t+i\epsilon)+\alpha\int\nu(\mathrm{d}\lambda)\frac{\phi''(\lambda)}{\alpha+\phi''(\lambda)g}=0. \tag{7.24}$$

So heuristically we can write for small enough ϵ:

$$\kappa_{\alpha,\phi}(\nu,t)=\operatorname*{extr}_{g\in\mathbb{C}_+}\Big\{-\ln|g|-tg_r+\epsilon g_i+\alpha\int\nu(\mathrm{d}\lambda)\ln|\alpha+\phi''(\lambda)g|-1-\alpha\ln\alpha\Big\}, \tag{7.25}$$

with $g=g_r+ig_i$ (in practice one considers the two variables g_r and g_i to find the extremum).

Heuristic solution to Theorem 7.5

We start from the result of Theorem 7.5. For a function f, we write $\mathbb{E}_\nu[f(X)]\equiv\int\nu(\mathrm{d}t)f(t)$. We introduce Lagrange multipliers to fix the conditions $\mathbb{E}_\nu[\phi(X)]\in B$, and we fix the values of $\mathbb{E}_\nu[\phi'(X)^2]$ and $\mathbb{E}_\nu[X\phi'(X)]$. We obtain:

$$\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)=\sup_{\substack{l\in B\\ \nu\in\mathcal{M}_1^+(\mathbb{R})}}\operatorname*{extr}_{\lambda_0,\lambda_1,\lambda_2}\sup_{A,t}\Big[\frac{1+\ln\alpha}{2}-\frac{1}{2}\ln A+\lambda_0 l+\lambda_1 A+\lambda_2 t\\ +\kappa_{\alpha,\nu}(\nu,t)-\alpha D_{\mathrm{KL}}(\nu|\mu_G)-\lambda_0\mathbb{E}_\nu[\phi(X)]-\lambda_1\mathbb{E}_\nu[\phi'(X)^2]-\lambda_2\mathbb{E}_\nu[X\phi'(X)]\Big].$$

The supremum over ν is now unconstrained over the set $\mathcal{M}_1^+(\mathbb{R})$ of probability distributions. We now make use of eq. (7.25) to write, with $2K(\alpha)\equiv-1+\ln\alpha-2\alpha\ln\alpha$ and a small $\epsilon>0$:

$$\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)=\sup_{\substack{l\in B\\ \nu\in\mathcal{M}_1^+(\mathbb{R})}}\operatorname*{extr}_{\substack{\{\lambda_i\},A,t\\ g\in\mathbb{C}_+}}\Big\{K(\alpha)-\frac{1}{2}\ln A+\lambda_0 l+\lambda_1 A+\lambda_2 t\\ -\ln|g|-t\mathrm{Re}[g]+\epsilon\mathrm{Im}[g]+\alpha\int\nu(\mathrm{d}\lambda)\ln|\alpha+\phi''(\lambda)g|-\alpha D_{\mathrm{KL}}(\nu|\mu_G)-\lambda_0\mathbb{E}_\nu[\phi(X)]\\ -\lambda_1\mathbb{E}_\nu[\phi'(X)^2]-\lambda_2\mathbb{E}_\nu[X\phi'(X)]\Big\}.$$

For any function $F:\mathbb{R}\to\mathbb{R}$, the maximum $\sup_\nu\{\mathbb{E}_\nu[F(X)]-\alpha D_{\mathrm{KL}}(\nu|\mu_G)\}$ is attained in ν^* with density proportional to $e^{-x^2/2+F(x)/\alpha}$, which is exactly the *Gibbs measure* of statistical physics, see Section 1.5. This gives ($\mathcal{D}$ is the standard Gaussian measure on $\mathbb{R}$):

$$\sup_{\nu\in\mathcal{M}_1^+(\mathbb{R})}\{\mathbb{E}_\nu[F(X)]-\alpha D_{\mathrm{KL}}(\nu|\mu_G)\}=\alpha\ln\int_{\mathbb{R}}\mathcal{D}x\,e^{F(x)/\alpha}.$$

Plugging this into our previous equation for the annealed complexity yields:

$$\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)=\sup_{l\in B}\operatorname*{extr}_{\substack{\{\lambda_i\},A,t\\ g\in\mathbb{C}_+}}\Big\{K(\alpha)+\lambda_0 l+\lambda_1 A+\lambda_2 t-\frac{\ln A}{2}-\ln|g|-t\mathrm{Re}[g]+\epsilon\mathrm{Im}[g]$$
$$+\alpha\ln\int_{\mathbb{R}}\mathcal{D}x\exp\Big\{-\frac{\lambda_0\phi(x)+\lambda_1\phi'(x)^2+\lambda_2 x\phi'(x)}{\alpha}+\ln|\alpha+\phi''(x)g|\Big\}\Big\}.$$

This can be further simplified, as the extrema over A,t are trivially solved and give the value of $\lambda_2=\mathrm{Re}[g]$ and $\lambda_1=(2A)^{-1}$. Thus we obtain:

$$\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)=\sup_{l\in B}\operatorname*{extr}_{\substack{\{\lambda_0,\lambda_1\}\\ g\in\mathbb{C}_+}}\Big\{K(\alpha)+\lambda_0 l+\frac{1+\ln 2}{2}+\frac{1}{2}\ln\lambda_1-\ln|g|+\epsilon\mathrm{Im}[g]$$
$$+\alpha\ln\int_{\mathbb{R}}\mathcal{D}x\exp\Big\{-\frac{\lambda_0\phi(x)+\lambda_1\phi'(x)^2+\mathrm{Re}[g]x\phi'(x)}{\alpha}+\ln|\alpha+\phi''(x)g|\Big\}\Big\}. \tag{7.26}$$

Let us now denote the measure:

$$\langle\cdots\rangle_{\lambda_0,\lambda_1,g}\equiv\frac{\int_{\mathbb{R}}\mathcal{D}x(\cdots)\exp\{-\alpha^{-1}[\lambda_0\phi(x)+\lambda_1\phi'(x)^2+\mathrm{Re}[g]x\phi'(x)]+\ln|\alpha+\phi''(x)g|\}}{\int_{\mathbb{R}}\mathcal{D}x\exp\{-\alpha^{-1}[\lambda_0\phi(x)+\lambda_1\phi'(x)^2+\mathrm{Re}[g]x\phi'(x)]+\ln|\alpha+\phi''(x)g|\}},$$

then the fixed point equations corresponding to the formula of eq. (7.26) can be written as:

$$\begin{cases} l & =\langle\phi(x)\rangle_{\lambda_0,\lambda_1,g},\\ \dfrac{1}{2\lambda_1} & =\langle\phi'(x)^2\rangle_{\lambda_0,\lambda_1,g},\\ -\dfrac{\mathrm{Re}[g]}{|g|^2} & =\Big\langle x\phi'(x)-\dfrac{\alpha\phi''(x)(\alpha+\phi''(x)\mathrm{Re}[g])}{|\alpha+\phi''(x)g|^2}\Big\rangle_{\lambda_0,\lambda_1,g},\\ \epsilon-\dfrac{\mathrm{Im}[g]}{|g|^2} & =-\Big\langle\dfrac{\alpha\phi''(x)^2\mathrm{Im}[g]}{|\alpha+\phi''(x)g|^2}\Big\rangle_{\lambda_0,\lambda_1,g}. \end{cases} \tag{7.27}$$

These equations are to be iterated over λ_0,λ_1,g, and l (while enforcing the constraint $l\in B$). From experience, the best procedure is to start from the solution of the unconstrained problem (without any constraint on the loss value), before smoothly following the solution while adding the constraint. In the case of $L_2(\mathbf{x})$, one would follow a similar procedure.

7.5 The quenched complexity and the replica method

In this section we detail the principle of the quenched calculation that gives rise to Results 7.1 and 7.2. For the sake of the presentation we restrict to Result 7.1, while Result 7.2 will be discussed in Appendix E.2.5. As the very basis of this calculation is non-rigorous we present this calculation in a fashion closer to theoretical physics standards, differently from Section 7.3 which was written in a more mathematical convention. Many technicalities will be postponed to Appendix E.1.

7.5.1 Computing the p-th moment

If needed, the reader can refer to Section 1.3.1 for an introduction to the replica method, one of the most important heuristic tools of theoretical statistical physics, that we already used extensively throughout this thesis. The first step of the method is to compute the integer

moments of the observable, here the number of critical points. Let $B \subseteq \mathbb{R}$ an open interval. The Kac-Rice formula can then be stated for the p-th moment of the complexity [AW09b, AT09]:

$$\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)^p = \Big[\prod_{a=1}^{p}\int_{\mathbb{S}^{n-1}}\mu_{\mathbb{S}}(\mathrm{d}\mathbf{x}^a)\Big]\mathbb{1}[\{L_1(\mathbf{x}^a)\in B\}_{a=1}^p]\,\varphi_{\{\mathrm{grad}\,L_1(\mathbf{x}^a)\}_{a=1}^p}(0)$$
$$\times\,\mathbb{E}\Big[\prod_{a=1}^{p}\det\mathrm{Hess}\,L_1(\mathbf{x}^a)\Big|\{\mathrm{grad}\,L_1(\mathbf{x}^a)\}_{a=1}^p=0\Big].$$

Here, $\varphi_{\{\mathrm{grad}\,L_1(\mathbf{x}^a)\}_{a=1}^p}(0)$ represents the joint density of the p gradients, taken at 0. Note that the functions $\{L_1(\mathbf{x}^a)\}_{a=1}^p$ only depend on the parameters $y_\mu^a \equiv \boldsymbol{\xi}_\mu\cdot\mathbf{x}^a$, so we will abusively write $L_1(\mathbf{y}^a)\equiv L_1(\mathbf{x}^a)$. Proceeding as in the annealed case, we can rewrite the expectations by conditioning over $\{\mathbf{y}^a\}_{a=1}^p$:

$$\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)^p = \Big[\prod_{a=1}^{p}\int\mu_{\mathbb{S}}(\mathrm{d}\mathbf{x}^a)\Big]\mathbb{E}_{\{\mathbf{y}^a\}}\Big\{\mathbb{1}[\{L_1(\mathbf{y}^a)\in B\}_{a=1}^p]\,\varphi_{\{\mathrm{grad}\,L_1(\mathbf{x}^a)\}_{a=1}^p|\{\mathbf{y}^a\}}(0)$$
$$\times\,\mathbb{E}\Big[\prod_{a=1}^{p}|\det\mathrm{Hess}\,L_1(\mathbf{x}^a)|\Big|\{\mathrm{grad}\,L_1(\mathbf{x}^a)=0,\mathbf{y}^a\}_{a=1}^p\Big]\Big\}. \tag{7.28}$$

The gradient and Hessian at $\mathbf{x}^a$ live in the tangent plane to the sphere at $\mathbf{x}^a$, identified with $\mathbb{R}^{n-1}$. Note that the $\{y_\mu^a\}$ are Gaussian variables with zero mean and covariance $\mathbb{E}[y_\mu^a y_\nu^b]=\delta_{\mu\nu}q_{ab}$, with $q_{ab}\equiv\mathbf{x}^a\cdot\mathbf{x}^b$ the "overlap" between replicas a and b. We introduce the variables $\{q_{ab}\}$ via delta functions in eq. (7.28):

$$\mathbb{E}\,\mathrm{Crit}_{n,L_1}(B)^p = \Big[\prod_{a=1}^{p}\int\mu_{\mathbb{S}}(\mathrm{d}\mathbf{x}^a)\Big]\Big[\prod_{a<b}\int\mathrm{d}q_{ab}\delta(q_{ab}-\mathbf{x}^a\cdot\mathbf{x}^b)\Big]\mathbb{E}_{\{\mathbf{y}^a\}}\Big\{\mathbb{1}[\{L_1(\mathbf{y}^a)\in B\}_{a=1}^p]$$
$$\varphi_{\{\mathrm{grad}\,L_1(\mathbf{x}^a)\}_{a=1}^p|\{\mathbf{y}^a\}}(0)\mathbb{E}\Big[\prod_{a=1}^{p}|\det\mathrm{Hess}\,L_1(\mathbf{x}^a)|\Big|\{\mathrm{grad}\,L_1(\mathbf{x}^a)=0,\mathbf{y}^a\}_{a=1}^p\Big]\Big\}. \tag{7.29}$$

Since we fixed the $\{q_{ab}\}$, the distribution of the $\{\mathbf{y}^a\}$ is fixed, as well as the joint distribution of the loss, gradients and Hessians, as we will explicit in the following. As the number of overlap variables is $p(p-1)/2=\mathcal{O}_n(1)$, we can apply Laplace's method over the variables $\{q_{ab}\}$ in the thermodynamic limit. As we explained in Section 1.3.1, we can make a *replica-symmetric* ansatz: It amounts to assume that, once Laplace's method is performed, the extremizing $\{q_{ab}\}$ satisfy $q_{aa}=1$, $q_{ab}=q$ for $a\neq b$. Assuming this structure of the overlap matrix allows to extend the expression of the moments to arbitrary non-integer p, and then to take the $p\downarrow 0$ limit as needed in the replica method. Note that while the RS structure is correct in many cases, it is also a very good approximation in others where replica symmetry is broken, which gives relevance to our RS calculation even in this case. We now detail the computation of the three factors of eq. (7.29), and their limit as $p\downarrow 0$ and $n\to\infty$.

The phase volume factor

Let us first compute the phase space factor in eq. (7.29). More precisely, the term:

$$\Big[\prod_{a=1}^{p}\int\mu_{\mathbb{S}}(\mathrm{d}\mathbf{x}^a)\Big]\Big[\prod_{a<b}\delta(q_{ab}-\mathbf{x}^a\cdot\mathbf{x}^b)\Big] = n^{-\frac{p(p-1)}{2}}\prod_{a=1}^{p}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{x}^a\prod_{a\le b}\delta(nq_{ab}-n\mathbf{x}^a\cdot\mathbf{x}^b),$$

in which we defined $q_{aa} = 1$. As we detail in Appendix E.1.1 we reach, when $p \downarrow 0$ and $n \to \infty$:

$$\frac{1}{np} \ln \prod_{a=1}^{p} \int_{\mathbb{R}^n} \mathrm{d}\mathbf{x}^a \prod_{a \le b} \delta(n q_{ab} - n \mathbf{x}^a \cdot \mathbf{x}^b) \simeq \frac{1}{2} \ln \frac{2\pi}{n} + \frac{1}{2}\Big[\frac{1}{1-q} + \ln(1-q)\Big]. \tag{7.30}$$

The joint density of the gradients

We will now compute the joint density of the gradients at $\{\mathbf{x}^a\}$, conditioned on the values of $\{\mathbf{y}^a\}$. This calculation is an extension of Sections V.C and V.E of [RBABC19]. We consider two vectors $\mathbf{x}^a$ and $\mathbf{x}^b$ of overlap $q_{ab} = q$. It is easy to see that $\mathbb{E}[\mathrm{grad}\, L(\mathbf{x}^a) | \{\mathbf{y}^b\}_{b=1}^p] = 0$ from eq. (7.14b), so we will focus on the covariance matrix $\mathbb{E}[\mathrm{grad}\, L(\mathbf{x}^a) \mathrm{grad}\, L(\mathbf{x}^b)^\intercal | \{\mathbf{y}^c\}_{c=1}^p]$. After some calculations detailed in Appendix E.1.2 we get the gradient density at leading exponential order:

$$\begin{aligned} \varphi_{\{\mathrm{grad}\, L_1(\mathbf{x}^a)\}_{a=1}^p | \{\mathbf{y}^a\}}(0) \simeq \prod_{a \neq b} \delta\Bigg[\frac{1}{m} \sum_{\mu=1}^{m} \phi'(y_\mu^a)\Big(z_p(q) y_\mu^a + f_p^0(q) y_\mu^b + f_p(q) \sum_{c(\neq a,b)} y_\mu^c\Big)\Bigg] \\ \times \exp\left\{\frac{np}{2} \ln \frac{m}{2\pi} - \frac{n}{2} \ln \det \left[\Big(\frac{1}{m} \sum_{\mu=1}^{m} \phi'(y_\mu^a)\phi'(y_\mu^b)\Big)_{1 \le a,b \le p}\right]\right\}, \end{aligned} \tag{7.31}$$

in which the auxiliary functions $(z_p(q), f_p(q), f_p^0(q))$ are explicit, and defined in eq. (E.3).

Factorization of the mean product of determinants

The argument of this section is very close to Section V.F of [RBABC19]. We consider the term:

$$\mathbb{E}\Big[\prod_{a=1}^{p} |\det \mathrm{Hess}\, L_1(\mathbf{x}^a)| \Big| \{\mathrm{grad}\, L_1(\mathbf{x}^a) = 0, \mathbf{y}^a\}_{a=1}^p\Big]. \tag{7.32}$$

We make two important remarks, which are straightforward transpositions of the arguments of [RBABC19] to our problem, and we refer to this work for more extensive physical justifications.

- The conditioning over the gradients being zero, similarly to what we showed in the annealed calculation, only gives a finite-rank change to the Hessians $\mathrm{Hess}\, L_1(\mathbf{x}^a)$ and thus does not modify the behavior of the determinant in the scale $e^{\Theta(n)}$. In this scale, the statistics of the p matrices $\{\mathrm{Hess}\, L_1(\mathbf{x}^a)\}_{a=1}^p$ are identical.
- For each a, the spectral measure of $\mathrm{Hess}\, L_1(\mathbf{x}^a)$ concentrates at a rate at least $n^{1+\epsilon}$ for a small enough $\epsilon > 0$ (we expect that the actual rate is n^2). This argument is very similar to what we used to prove Lemma 7.10: this implies that at the order $e^{\Theta(n)}$ the expectation value factorizes over the replicas, and we can assume all the Hessians to be independent.

Before stating the consequences of such remarks, we give some definitions:

- $\mu_{\mathrm{G},q}$ is the Gaussian probability measure on $\mathbb{R}^p$ with zero mean and covariance $\mathbb{E}[X_a X_b] = (1-q)\delta_{ab} + q$. Note that $\{\mathbf{y}_\mu\}_{\mu=1}^m$ are i.i.d. variables distributed according to $\mu_{\mathrm{G},q}$.
- We define $\nu_{\mathbf{y}}$ as the empirical measure of $(\mathbf{y}_1, \cdots, \mathbf{y}_m)$, that is $\nu_{\mathbf{y}} \equiv (1/m) \sum_\mu \delta_{\mathbf{y}_\mu}$. For every a, we denote $\nu_{\mathbf{y}}^a$ its marginal distribution: $\nu_{\mathbf{y}}^a(\mathrm{d}\lambda^a) \equiv \int \prod_{b(\neq a)} \nu_{\mathbf{y}}(\mathrm{d}\lambda)$. Then $\nu_{\mathbf{y}}^a$ is also the empirical distribution of $(y_\mu^a)_{\mu=1}^m$.

Our remarks show that we can use the results of the annealed calculation, and that the expectation of the determinants factorizes at leading exponential order:

$$\mathbb{E}\Big[\prod_{a=1}^{p} |\det \mathrm{Hess}\, L_1(\mathbf{x}^a)| \Big| \{\mathrm{grad}\, L_1(\mathbf{x}^a)\}_{a=1}^{p} = 0,\ \{\mathbf{y}^a\}\Big] \simeq e^{n\sum_{a=1}^{p} \kappa_{\alpha,\phi}\left(\nu_{\mathbf{y}}^a, t_\phi(\nu_{\mathbf{y}}^a)\right)}. \tag{7.33}$$

As for the annealed case, the concentration behind eq. (7.33) has recently been shown for a very large of class of random matrices by collaborators [BABM21a] (eq. (1.2) of this work is precisely equivalent to our eq. (7.33)).

7.5.2 Decoupling replicas and the $p \downarrow 0$ limit

We can then apply Sanov's theorem 1.9 to the empirical measure $\nu_{\mathbf{y}} \in \mathcal{M}_1^+(\mathbb{R}^p)$. Recall that we have constraints on this measure by the density of the gradient and the fixation of the energy level. More precisely, we denote $\mathcal{M}_\phi^{(p)}(q, B)$ the set of probability measures on $\mathbb{R}^p$ that satisfy the following:

$$\begin{cases} \displaystyle\int \nu(\mathrm{d}\lambda)\, \phi(\lambda^a) \in B & \forall 1 \le a \le p, \qquad (7.34a) \\ \displaystyle\int \nu(\mathrm{d}\lambda)\, \phi'(\lambda^a) \Big[z_p(q)\lambda^a + f_p^0(q)\lambda^b + f_p(q) \sum_{c(\neq a,b)} \lambda^c \Big] = 0 & \forall 1 \le a \neq b \le p. \qquad (7.34b) \end{cases}$$

Recall that the functions $(z_p(q), f_p^0(q), f_p(q))$ are defined in eq. (E.3). Leveraging from the results of eqs. (7.30), (7.31) and (7.33), we obtain from Sanov's theorem 1.9 and Varadhan's lemma 1.10:

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\left[\mathrm{Crit}_{n,L_1}(B)^p\right] = \frac{p}{2} \ln \alpha + \sup_{q\in(0,1)} \sup_{\nu \in \mathcal{M}_\phi^{(p)}(q,B)} \Big[\frac{p}{2}\Big(\frac{1}{1-q} + \ln(1-q)\Big) \tag{7.35}$$

$$- \frac{1}{2} \ln \det \Big[\Big(\int \nu(\mathrm{d}\lambda) \phi'(\lambda^a)\phi'(\lambda^b) \Big)_{1\le a,b\le p} \Big] + \sum_{a=1}^{p} \kappa_{\alpha,\phi}(\nu^a, t_\phi(\nu^a)) - \alpha D_{\mathrm{KL}}(\nu | \mu_{\mathrm{G},q}) \Big].$$

Recall that ν^a is the marginal distribution of ν for the variable λ^a. We can then decouple the replicas under an assumption on the measure ν that amounts for replica symmetry. We stress that this replica symmetric assumption in the Kac-Rice calculation actually corresponds to a *1-step replica symmetry breaking (1RSB) structure of the zero-temperature Gibbs measure*, that is an exponential number of single-point metastable states that all have the same two-point overlap. While possibly not exact, this assumption should already yield a good approximation to the landscape, and could be analytically checked by studying the stability of the replica-symmetric ansatz within replica theory. This allows to take subsequently the $p \downarrow 0$ limit, and after some simplifications, we reach from eq. (7.35) the expression of Result 7.1. These steps are fairly technical, and are postponed to Appendix E.1.3.

Discussion on Chapter 7

In this chapter, we obtained analytical results for the annealed and quenched complexities of statistical models with non-Gaussian loss functions arising in generalized linear estimation and simple models of glasses and neural networks. Our method is versatile and can be easily extended to other cases, as we will discuss below.

Our results allow for a complete characterization of the empirical loss landscapes of generalized linear models. The main issue ahead is determining for which class of functions ϕ and in

which regimes (e.g. values of α), the annealed and quenched complexities become positive, i.e. when the associated landscape is rough, and if they differ, i.e. if the number of critical points is concentrating as the dimension grows large. These investigations will surely require a systematic and extensive numerical evaluation of our results, and will allow to study the connection between landscape properties and dynamics induced by local algorithms. In particular, it will shed light on the relationship between the roughness of the empirical loss landscape and the existence of "hard" phases in the learning of generalized linear models [BKM+19]. It will also provide an interesting benchmark for obtaining the algorithmic thresholds of gradient descent (and variants) only through the knowledge of the landscape properties [SMBC+19, SMBC+20b, BAGJ21]. Based on ongoing works, we can for instance conjecture the existence of a rough landscape for small enough α in phase retrieval [LSL19] and retarded learning [EVdB01]. Addressing these questions is an exciting direction of research, and a natural follow-up to this chapter that is under investigation.

Counting the minima – Another important extension of this chapter consists in counting the critical points of a fixed index (i.e. with a fixed number of negative directions in the spectrum of the Hessian). This would provide additional interesting information, in particular it would allow to differentiate local minima from the other critical points of the landscape, as we did for the p-spin model in Section 7.1. As we mentioned in this spin glass calculation, such a counting would require to understand the large deviations of the eigenvalues of the Hessian arising for generalized linear models, i.e. random matrices of the type of eq. (7.13). Such a random matrix problem is hard, but in Chapter 8 we present a technique that builds on recent developments and that allows to compute the rate function of the large deviations of the smallest eigenvalue. This allows then to use our Kac-Rice calculation to count solely minima of the loss!

Generalization to other models – Our calculations, both annealed and quenched, can be generalized straightforwardly to many other loss functions and models. As is clear for instance in the annealed computation of Section 7.3, the key features that must be present are:

1. A Gaussian distribution of the data $\boldsymbol{\xi}_\mu$.
2. A loss function $L(\mathbf{x})$ that only depends on the data samples $\boldsymbol{\xi}_\mu$ via their projection over a few vectors (e.g. $\mathbf{x}$ for $L_1(\mathbf{x})$ and $\mathbf{x}, \mathbf{x}^\star$ for $L_2(\mathbf{x})$).

We give thereafter three examples of models, that can be found in [EVdB01, MBM18], and in which our calculations can be easily performed:

Model 7.1 (*Binary linear classification*)

Consider $n, m \geq 1$ such that $m/n \to \alpha > 1$. Let $\sigma : \mathbb{R} \to [0,1]$ a smooth threshold function. We are given m samples $(y_\mu, \mathbf{x}_\mu)_{\mu=1}^m$ with $y_\mu \in \{0,1\}$ and $\mathbf{x}_\mu \in \mathbb{R}^n$. The elements of $(y_\mu)_{\mu=1}^m$ are generated according to $\mathbb{P}(Y_\mu = 1 | \mathbf{X}_\mu = \mathbf{x}) = \sigma(\boldsymbol{\theta}_0 \cdot \mathbf{x})$, and $\mathbf{x}_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_n)$. We want to learn the vector $\boldsymbol{\theta}_0 \in \mathbb{S}^{n-1}$ by minimizing the following loss function:

$$L(\boldsymbol{\theta}) \equiv \frac{1}{2m} \sum_{\mu=1}^{m} [y_\mu - \sigma(\boldsymbol{\theta} \cdot \mathbf{x}_\mu)]^2, \qquad \boldsymbol{\theta} \in \mathbb{S}^{n-1}. \tag{7.36}$$

Model 7.2 (*Mixture of two Gaussians*)

Consider $n, m \geq 1$ such that $m/n \to \alpha > 1$. We are given m samples $\mathbf{y}_\mu \in \mathbb{R}^n$, generated as $\mathbf{y}_\mu \overset{\text{i.i.d.}}{\sim} \sum_{a=1}^2 p_a \mathcal{N}(\boldsymbol{\theta}_a^0, \mathrm{I}_n)$. The proportions p_1, p_2 are known, and we wish to recover $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$ by minimizing the maximum-likelihood estimator:

$$L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \equiv -\frac{1}{m}\sum_{\mu=1}^m \ln\Big[\frac{1}{2}\sum_{a=1,2}\frac{1}{\sqrt{2\pi}^n}\exp\Big\{-\frac{1}{2}\|\mathbf{y}_\mu - \boldsymbol{\theta}_a\|^2\Big\}\Big]. \tag{7.37}$$

Model 7.3 (*Simple unsupervised learning model*)

Consider $n, m \geq 1$ such that $m/n \to \alpha > 1$. Let $\phi : \mathbb{R} \to \mathbb{R}$ a smooth activation function, $V : \mathbb{R} \to \mathbb{R}_+$ a "potential", and $\mathbf{x}^0 \in \mathbb{S}^{n-1}$ a fixed vector. We assume that we are given i.i.d. data samples $\{\boldsymbol{\xi}_\mu\}_{\mu=1}^m \in \mathbb{R}^n$ distributed such that their projection on $\mathbf{x}^0$ has a probability density $P(\boldsymbol{\xi}_\mu \cdot \mathbf{x} = h) \propto e^{-\frac{1}{2}h^2 - V(h)}$, and the other coordinates of $\boldsymbol{\xi}_\mu$ are i.i.d. standard Gaussian variables. We wish to recover the vector $\mathbf{x}^0$ by minimizing:

$$L(\mathbf{x}) \equiv \frac{1}{m}\sum_{\mu=1}^m \phi(\boldsymbol{\xi}_\mu \cdot \mathbf{x}), \qquad \mathbf{x} \in \mathbb{S}^{n-1}. \tag{7.38}$$

For each of these three models one can easily replicate the annealed and quenched calculations of Sections 7.3 and 7.5, under suitable technical hypotheses. As a final note, it is however an open problem to generalize our methods to neural network models with many nodes and hidden layers; the random matrix analysis of the Hessian in this case is a particularly exciting challenge.

A note on non-spherical priors – It is clear from the calculation of Appendix E.1 (particularly Section E.1.1) that we can also generalize our techniques (at least on the heuristic level) to non-spherical prior distributions on the vectors $\mathbf{x}$. The most natural hypothesis that allows the computation to be generalized is that the prior distribution takes the decoupled form $P(\mathrm{d}\mathbf{x}) = \prod_i P(\mathrm{d}x_i)$.

Going further: discrete systems – In a series of recent fascinating papers [Sub21, Mon21, AMS20, AM20, Sel21], the authors developed polynomial-time algorithms that can provably optimize a large class of discrete and continuous disordered systems (such as the p-spin on the hypercube). However, the very notion of "critical points" is generally not well-defined in discrete problems, thus preventing from using the Kac-Rice method as we did in this chapter. Looking for topological invariants of the landscape to characterize the optimal algorithmic performance in discrete models is therefore an open and exciting research direction. First explorations have been performed using the Kac-Rice formula on the corresponding continuous TAP landscape (cf. Section 1.3.2), cf. e.g. for Ising spins [CPS21, FMM21].

Erratum to the proof – Finally, note that the proof of Theorem 7.5 that we presented here has a minor modification with respect to the one published in [MBAB20], as we uncovered a small missing step in the proof. This addition will soon be present in the published version of this work as well.

Chapter 8

An excursion to large deviations in random matrix theory

"On voit par cet Essai, que la théorie des probabilités n'est au fond, que le bon sens réduit au calcul : elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte."

Pierre-Simon de Laplace, Essai philosophique sur les probabilités (1814).

Disclaimer – In this chapter we present an analytical technique to compute the probability of rare events in which the largest eigenvalue of a large class of random matrices, known as *generalized sample covariance matrices*, is atypically large (i.e. the right tail of its large deviations). The results also transfer to the left tail of the large deviations of the smallest eigenvalue. In particular, these include the Hessian of the loss of the empirical risks L_1 and L_2 of eqs. (7.9),(7.10). As we mentioned in Section 7.1, the calculation of these large deviations is primarily motivated by the ability to compute the complexity of local minima in the class of inference models studied in Chapter 7. Moreover we will detail below other strong motivations arising from theoretical statistics to study these large deviations. This chapter is based on the published work [Mai21]. While it revolves around subjects very close to Chapter 7, it is written in a style much closer to the theoretical physics literature; in particular, we will often not use theorems to designate our results. However, given the nature of the technique we use, which originated in the mathematics literature, we expect that a rigorous proof would follow exactly the steps taken here.

8.1 Why the large deviations of the eigenvalues?

8.1.1 The landscape of generalized linear models and variants

The first motivation behind our study naturally comes from our results of Chapter 7, and the Kac-Rice formalism developed there. Recall that this chapter focused on the study of the complexity of the empirical risk landscape of generalized linear models. One of the most important extensions of this chapter would be to restrict the counting of critical points to local minima, which are more representative of the actual roughness of the landscape from the point of view of local optimization algorithms.

As the Kac-Rice formula turns the counting of the complexity into the random matrix analysis of the Hessian, we saw that conditioning a critical point to be a local minimum naturally requires to understanding very rare events in which the smallest eigenvalue of the Hessian has extremely atypical value, i.e. its *large deviations* (as defined in Section 1.5.2). Exploring precisely these inference landscapes is an important open problem for the disordered systems and statistical learning communities, as the traditional methods have been limited to simpler Gaussian models (see e.g. [MBM18, SMBC$^+$19, SMKUZ19, RBABC19] and many other references given in

Chapter 7), and the large deviations results of the present chapter are an important step in this exciting direction.

8.1.2 PCA for correlated data

On the other hand, a textbook example of the interplay between theoretical physics and statistics (that fuels this thesis) is principal components analysis (PCA), a statistical estimation method based on random matrix theory, and applied in fields as diverse as image compression [DF07, ZDZS10, AW10, Fuk13], neurosciences [JFHK94, BBVS00], genetics [RPP08], or finance [BP00]. In Chapter 5 we studied variants of this problem through the prism of generative models.

To fix our ideas, let $\mathbf{X} \in \mathbb{R}^{m \times n}$ be the data matrix, whose columns $\{\mathbf{x}_i\}_{i=1}^n$ are observations independently drawn from a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$. PCA aims at discovering a "principal component" eigenspace of the covariance matrix $\mathbf{\Gamma}$ by studying the largest eigenvalue of the *sample covariance matrix* $\mathbf{C}_n \equiv \sum_i \mathbf{x}_i \mathbf{x}_i^\intercal / n$: indeed, a strong outlier eigenvalue in Γ typically induces a corresponding outlier in $\mathbf{C}_n$ [EJ76, BBAP05, BGN11] (see as well Chapter 5).

Pioneering physics works addressed the general question "How good is PCA ?" [DM06, MV09]. Precisely, they wished to understand if an outlier can appear in $\mathbf{C}_n$ even if there is no structure to uncover in $\mathbf{\Gamma}$: this "null hypothesis" provides a way to gauge the significance of results obtained on a real-world dataset. As we defined in Section 1.5.2, such atypical events are known as *large deviations*, and the mentioned works, as well as subsequent ones, had to restrict to *uncorrelated* data, in which $\mathbf{\Gamma}$ is the identity matrix (or a finite-rank perturbation of it) [DM06, Mai07, VMB07, MV09, MS14, BG20]. Realistic data (e.g. a natural image) indeed contain non-trivial correlations that the Coulomb gas analysis used in [DM06, MV09] is not equipped to handle. While data structure is a key ingredient of learning and inference (see [Zde20] and our discussion in Chapter 5), probing the statistical significance of PCA on correlated data remained an open question. In this regard, the present chapter addresses and solves this long-lasting problem for *arbitrary* $\mathbf{\Gamma}$, i.e. PCA with correlated data.

8.1.3 Organization of the chapter

We begin in Section 8.2 by defining precisely the class of matrices we consider, and we recall some known results on the behavior of its asymptotic spectrum. The main contribution of this chapter is Result 8.1, which gives the rate function of the large deviations of the largest eigenvalue of the class of matrices we consider. Our derivation is based on a *tilting* method, developed in a series of recent mathematical works [BG20, GH20, Hus20, BGH20, AGH21, McK21b]. This technique is more adaptable than a more traditional Coulomb gas analysis, as it does not require the joint probability of the eigenvalues of the matrix, which is unknown here. Moreover, the calculation does not rely on any heuristics, and we therefore expect it to be adaptable into mathematically rigorous statements. In Section 8.3 we numerically probe our results using an importance sampling Monte-Carlo approach, effectively simulating events with probability as small as 10^{-100}. Finally, Section 8.4 is devoted to the detailed derivation of the rate function, i.e. of Result 8.1.

8.2 Large deviations of extreme eigenvalues of generalized sample covariance matrices

8.2.1 Some formal definitions and assumptions

To state our main result, we need first to define some important mathematical concepts.

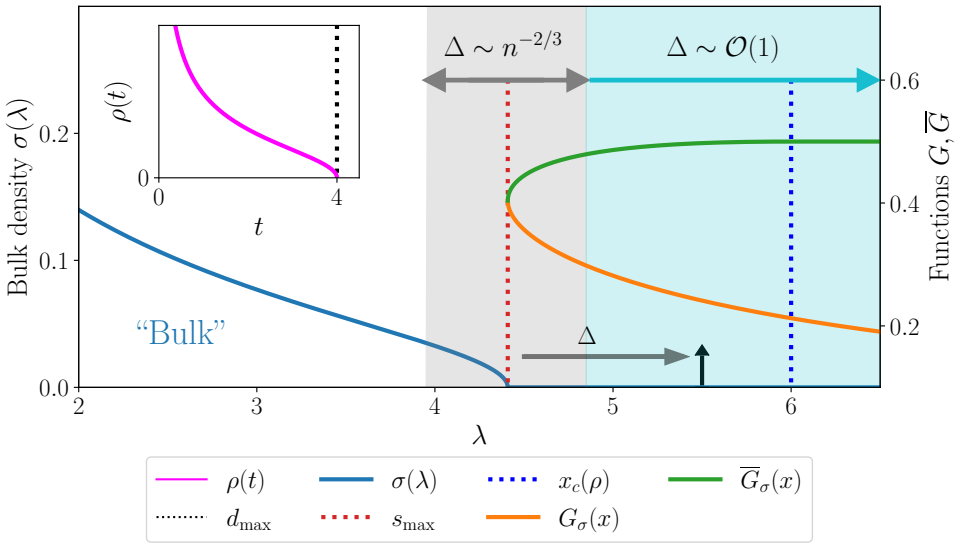


FIGURE 8.1: The bulk $\sigma(\lambda)$, and the functions $G_\sigma, \overline{G}_\sigma$ for $\alpha = 2$ and $\rho(t)$ the Marchenko-Pastur law with ratio 1. In the box, we plot $\rho(t)$ and the right edge $d_{\max}$ of its support. The black arrow is an outlier in the spectrum of $\mathbf{H}_n$, and Δ is the gap between this outlier and the bulk $\sigma(\lambda)$. As we will see, the rate function of the large deviations of $\lambda_{\max}(\mathbf{H}_n)$ is directly proportional to the area between G_σ and $\overline{G}_\sigma$.

Bulk density and Marchenko-Pastur equation

Letting $\mathbf{x}_i = \sqrt{\mathbf{\Gamma}}\mathbf{z}_i$ with $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathrm{I}_n)$, one can see that $\mathbf{C}_n$ has the same eigenvalues (up to possible zeros and a scaling factor) as $\mathbf{H}_n \equiv \mathbf{Z}^\intercal\mathbf{\Gamma}\mathbf{Z}/m$. The "bulk" of $\mathbf{H}_n$ – i.e. the large n limit of its eigenvalue density, or LSD – is denoted $\sigma(\lambda)$:

$$\int \mathrm{d}\lambda\, \sigma(\lambda)\, f(\lambda) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f[\lambda_i(\mathbf{H}_n)].$$

The LSD of $\mathbf{\Gamma}$ is in turn denoted $\rho(t)$. We illustrate these different quantities in Fig. 8.1, when $\rho(t)$ is the Marchenko-Pastur law with ratio one. The other functions pictured in Fig. 8.1 will be introduced later on.

Importantly, $\sigma(\lambda)$ can be analytically derived using the *Stieltjes transform* of random matrix theory, that we introduced in Section 1.5. Here we adopt a slightly different convention than in the rest of this thesis, namely we will consider:

$$G_\sigma(x) \equiv -\mathcal{S}_\sigma(x) = \int \mathrm{d}\lambda\, \frac{\sigma(\lambda)}{x - \lambda}.$$

Assuming that $m/n \to \alpha$, the Marchenko-Pastur equation [MP67] (Theorem 1.7) gives the inverse of G_σ:

$$G_\sigma^{-1}(\omega) = \frac{1}{\omega} + \alpha \int \mathrm{d}t \rho(t) \frac{t}{\alpha - t\omega}. \tag{8.1}$$

The bulk density $\sigma(\lambda)$ is then determined via the *Stieltjes-Perron* inversion formula (Theorem 1.5). In particular, the support of $\sigma(\lambda)$ and its right edge $s_{\max}$ can be computed (analytically or numerically) from eq. (8.1).

Generalized sample covariance matrices

By rotation invariance of $\mathbf{Z}$, one can diagonalize $\mathbf{\Gamma}$, i.e. assume $\mathbf{\Gamma} = \mathrm{Diag}(\{d_\mu\})$, with all $d_\mu \geq 0$, which implies that:

$$\mathbf{H}_n \equiv \frac{1}{m}\sum_{\mu=1}^{m} d_\mu \mathbf{z}_\mu \mathbf{z}_\mu^\dagger, \tag{8.2}$$

in which $\mathbf{z}_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathrm{I}_n)$. This form leads us to further extend the random matrix model we consider. More precisely, given a set of variables $d_\mu \in \mathbb{R}$ (*not necessarily positive*), we call matrices of the type of eq. (8.2) *generalized sample covariance matrices*. We can furthermore allow the $\mathbf{z}_\mu$ to be real or complex Gaussian variables, and recall that we consider the limit $n, m \to \infty$ with $m/n \to \alpha > 0$. The limit of the empirical distribution of $\{d_\mu\}_{\mu=1}^m$ is denoted $\rho(t)$, and we define $d_{\max} \in \mathbb{R} \cup \{+\infty\}$ the right edge of its support.

Importantly, the positivity (or negativity) of the matrix $\mathbf{H}_n$ is equivalent to the positivity (or negativity) of *all* d_μ. Coherently with our motivation, these matrices (with arbitrary $d_\mu \in \mathbb{R}$) contain, up to a shift, the Hessian matrix of eq. (7.13) arising in the computation of the complexity of the empirical risk of generalized linear models via the Kac-Rice formula.

Finally, we note that generalized sample covariance matrices are precisely the class of matrices covered by the generic form of the Marchenko-Pastur equation (Theorem 1.7).

The functions $G_\sigma, \overline{G}_\sigma$

By monotonicity arguments, it can easily be seen that the equation $G_\sigma^{-1}(\omega) = x$, with G_σ^{-1} given by eq. (8.1), can have more than one solution! More precisely, it can have exactly one other solution $\omega = \overline{G}_\sigma(x)$, sometimes referred to as the "second branch" of the Marchenko-Pastur equation. In Fig. 8.1, we show an example of functions $(G_\sigma, \overline{G}_\sigma)$. These functions satisfy the following properties[1]:

- G_σ is decreasing, while $\overline{G}_\sigma$ is non-decreasing.
- $G_\sigma(x) \sim_{x\to\infty} x^{-1}$.
- $\lim_{x\downarrow s_{\max}} G_\sigma(x) = \lim_{x\downarrow s_{\max}} \overline{G}_\sigma(x)$.
- For $d_{\max} \in (0, +\infty)$ (and therefore nonnegative $\mathbf{H}_n$), let:

  $$x_c(\rho) \equiv d_{\max}^2 G_\rho(d_{\max}) + (\alpha^{-1} - 1) d_{\max} \in (s_{\max}, +\infty]. \tag{8.3}$$

  Here, $G_\rho(z) \equiv -\mathcal{S}_\rho(z) = \int \mathrm{d}t\, \rho(t)/(z-t)$ is the (negative of the) Stieltjes transform of $\rho(t)$. If $x_c(\rho) < +\infty$, then the equation $G_\sigma^{-1}(\omega) = x$ has a single solution for $x > x_c(\rho)$, which is $G_\sigma(x)$. In this case, we define $\overline{G}_\sigma(x) \equiv \alpha/d_{\max}$ for $x \geq x_c(\rho)$. One can check that this ensures that $\overline{G}_\sigma(x)$ is continuous in $x_c(\rho)$: we will say that $\overline{G}_\sigma$ *saturates* at the point $x_c(\rho)$. An example of this saturation point is pictured in Fig. 8.1.
- If $\mathbf{H}_n$ is negative (i.e. $d_{\max} \leq 0$), then

  $$\lim_{x\uparrow 0} \overline{G}_\sigma(x) = +\infty.$$

  In this case, we set $\overline{G}_\sigma(x) = +\infty$ for $x \geq 0$.

[1] The derivation of these properties is straightforward and left to the reader.

8.2.2 Main result

From now on we will restrict to the study of $\lambda_{\max}(\mathbf{H}_n)$. Since we can always consider $d'_\mu = -d_\mu$, our analysis also applies to $\lambda_{\min}(\mathbf{H}_n)$, so that we do not lose any generality. As we know, the large deviations regime corresponds to *macroscopic* changes in $\lambda_{\max}(\mathbf{H}_n)$, which are exponentially rare, as opposed to the typical fluctuations, which are generically in the scale $n^{-2/3}$ for such random matrices [TW94, Joh01, DY20]: these two regimes are shown as cyan and grey regions in Fig. 8.1. Crucially, we assume

Hypothesis 8.1 (*"No outliers hypothesis"*)

Recall that we denote $\rho(t)$ the empirical distribution of $\{d_\mu\}_{\mu=1}^m$, and $d_{\max}$ the right edge of the support of $\rho(t)$. We assume that $d_{\max} < \infty$, and moreover:

$$\lim_{m\to\infty} \max_{1\leq\mu\leq m} d_\mu = d_{\max}$$

In other words, *there is no outlier in the list* $\{d_\mu\}$.

Importantly, Hypothesis 8.1 ensures that $\lambda_{\max}(\mathbf{H}_n)$ converges to the right edge $s_{\max}$ of the bulk $\sigma(\lambda)$. This implies that the set of $\{\mathbf{z}_\mu\}$ such that the spectrum of $\mathbf{H}_n$ has an outlier is very atypical under the Gaussian distribution. Let us now state our main result.

We adopt a standard notation, used e.g. in Chapters 2,6: we let $\beta \in \{1, 2\}$ for respectively real and complex $\mathbf{z}_\mu$, with the convention $\mathbb{E}|z|^2 = 1$ for a Gaussian standard random variable. We state the following result under all aforementioned hypotheses.

Result 8.1 (*Large deviations of the largest eigenvalue of $\boldsymbol{H}_n$*)

Recall that the variables $\{d_\mu\}$ are *given*. Under the randomness of $\{\mathbf{z}_\mu\}_{\mu=1}^m$, the law of $\lambda_{\max}(\mathbf{H}_n)$ satisfies a large deviation principle, in the scale n, with rate function $I(x)$ given by:

$$I(x) = \begin{cases} +\infty & \text{if } x < s_{\max} \\ \dfrac{\beta}{2}\displaystyle\int_{s_{\max}}^{x} [\overline{G}_\sigma(u) - G_\sigma(u)]\mathrm{d}u & \text{if } x \geq s_{\max}. \end{cases}$$

Result 8.1 is the main result of this chapter, and will be derived in Section 8.4. In Fig. 8.2, we show analytical computations of the rate function $I(x)$ for different α and $\rho(t)$. As we mentioned in the introduction of this chapter, we state this LDP as a result rather than a theorem, as its publication was made in a physics journal [Mai21]. As we will see however, we expect all arguments to transfer into a rigorous proof without significant changes. Let us now draw some first consequences of Result 8.1.

Consistency with previous results

Importantly, in the white Wishart case – i.e. $\rho(t) = \delta(t-1)$ – such large deviations were already analyzed in e.g. [MV09, BG20]. Our result should therefore be consistent with their findings. As detailed in Appendix E.3, Result 8.1 indeed reduces in this case to the previously-known expression:

$$I(x) = \frac{\alpha\beta}{2}\int_{\lambda_+(\alpha)}^{x} \frac{\sqrt{(u-\lambda_+(\alpha))(u-\lambda_-(\alpha))}}{u}\mathrm{d}u,$$

with $\lambda_+(\alpha) \equiv (1+\alpha^{-1/2})^2$, and $x \geq \lambda_+(\alpha)$.

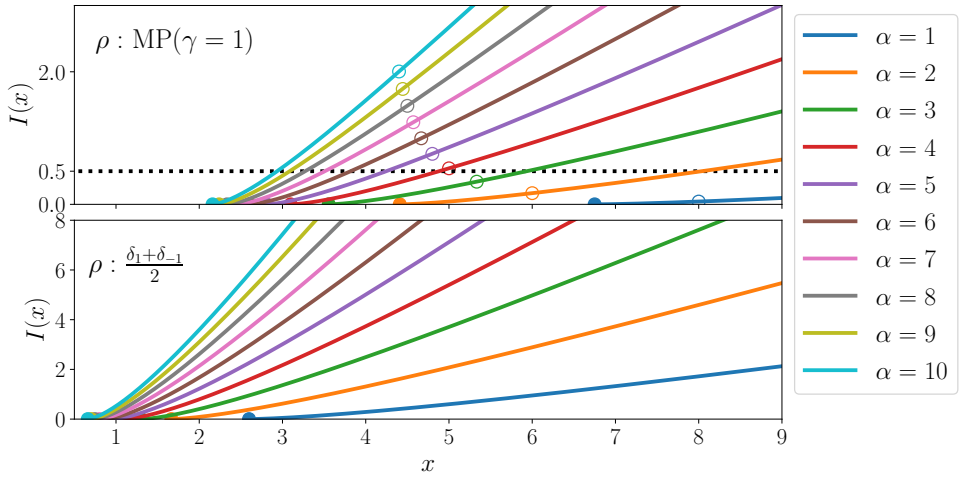


FIGURE 8.2: The rate function $I(x)$ for different values of α and two different distributions ρ, in the real case. Full dots depict the right edge $s_{\max}$ of the bulk, while empty dots (when present) correspond to the transition $x_c(\rho)$. We draw the dotted line $I(x) = 0.5$ in the top plot for later use.

A phase transition in the rate function

Let us describe a first notable consequence of Result 8.1. We assume that $d_{\max} > 0$ and that $x_c(\rho) < +\infty$. For instance, this is true if ρ is the Marchenko-Pastur law, as shown in Figs. 8.1 and 8.2.

We saw that in this case $\overline{G}_\sigma(x)$ saturates at $\alpha/d_{\max}$ for $x \geq x_c(\rho)$. It is in general not smooth at $x = x_c(\rho)$ and this singularity induces a *phase transition* in the rate function $I(x)$. The *order* of the transition – i.e. the order of the first discontinuous derivative of $I(x)$ – can be computed assuming the right tail of $\rho(t)$ behaves as

$$\rho(t) \sim_{t \uparrow d_{\max}} (d_{\max} - t)^\eta,$$

with $\eta > 0$, so that $x_c(\rho) < \infty$. When $\eta \geq 1$ and $1/2 \leq \eta < 1$ (e.g. the Marchenko-Pastur law, for which $\eta = 1/2$) we show that the transition is respectively of second and third order[2]. The details are given in Appendix E.4, and we conjecture generically the order of the transition to be $k+1$ if $1/k \leq \eta < 1/(k-1)$.

8.3 Monte-Carlo simulations

While Result 8.1 is in essence high-dimensional, this section is devoted to numerically probe its predictions. As we will see, even at moderately large n we are able to recover the rate function we computed analytically! While this regime is not directly relevant to the Kac-Rice computations of Chapter 7 (which are analytical results in the large n limit), it is the appropriate regime to investigate correlated data in PCA, which is our second main motivation for this study.

Importance sampling

Since we need $(1/m)\sum_\mu \delta(t - d_\mu)$ to be very close to $\rho(t)$ (and thus large enough m), we can not perform simple histograms of $\lambda_{\max}(\mathbf{H}_n)$, as e.g. in [MV09], since the large deviations probability decays exponentially in n. Instead, we will modify (or "tilt") the law of $\mathbf{z}$ so that it favors large

[2]Note that a similar argument based on the vanishing exponent of the density was already used in the literature, in the context of multi-critical matrix models [MS14].

deviations, a technique which is known as *importance sampling* [Buc13]. This powerful Monte-Carlo method allows to numerically access the tails of a given high-dimensional probability distribution and has been successfully applied to various problems across the physical sciences, from random graphs [EMH04] to simulations of the height distribution in the Kardar–Parisi–Zhang equation [HLDM+18], and random matrices [DM07, SIH10], as in this chapter. For a more exhaustive description of the applications of importance sampling in physics, we refer the reader to [HLDM+18].

Tilting the measure

Let us now detail our approach in detail. For the purpose of the presentation we will restrict to the real case $\beta = 1$, while all subsequent arguments can straightforwardly be adapted to the complex case. Recall that we denote $\mathcal{D}\mathbf{z} \equiv \mathrm{d}\mathbf{z}\, e^{-\|\mathbf{z}\|^2/2}/(2\pi)^{n/2}$ the standard Gaussian law. We will *tilt* this distribution by explicitly giving more weight to configurations having a larger $\lambda_{\max}(\mathbf{H}_n)$, making them more probable. More precisely, we aim at sampling from the distribution

$$P_t(\mathbf{z})\mathrm{d}\mathbf{z} \propto \mathcal{D}\mathbf{z}\, e^{nt\lambda_{\max}(\mathbf{H}_n)}. \tag{8.4}$$

For a given $t \geq 0$, Result 8.1 and Laplace's method imply that when sampling $\mathbf{z}$ under P_t, the largest eigenvalue of $\mathbf{H}_n$ concentrates on

$$x^\star(t) \equiv \operatorname*{arg\,min}_{x \geq s_{\max}} [tx - I(x)]. \tag{8.5}$$

One sees clearly now that sampling from the tilted distribution of eq. (8.4) gives information about the Legendre transform of the large deviations function $I(x)$.

The Monte-Carlo algorithm

We implement a classical Metropolis-Hastings algorithm in order to sample from P_t. The physical (given) parameters are n, m, ρ, t, and we generate i.i.d. samples $\{d_\mu\}_{\mu=1}^m$ from ρ. We pick two hyperparameters $\Delta, \beta_d > 0$ (we will later detail how to fine-tune them). We initialize $\{\mathbf{z}_\mu\}_{\mu=1}^m$ as standard Gaussian vectors, and we sample from the move proposal distribution $g(\mathbf{z}'|\mathbf{z})$ as follows:

(*i*) Pick a random index $\mu \in \{1, \cdots, m\}$ with probability $P(\mu) \propto e^{\beta_d d_\mu}$.

(*ii*) Draw a uniform vector $\mathbf{e} \in \mathbb{S}^{n-1}(\sqrt{n})$, and draw $L \geq 0$ from a truncated Gaussian distribution centered in 1 and with variance $\Delta > 0$. Let $\mathbf{z}'_\mu = \sqrt{L}\mathbf{e}$.

(*iii*) The new state is given by changing $\mathbf{z}_\mu \to \mathbf{z}'_\mu$.

We impose the detailed balance condition with stationary distribution $P_t(\mathbf{z})$ and move proposal distribution $g(\mathbf{z}'|\mathbf{z})$ in the MCMC. We measure the largest eigenvalue $\lambda_{\max}(\mathbf{H}_n)$, which we then compare to $x^\star(t)$. The parameters (β_d, Δ) are found to reduce greatly the equilibration time of the Markov chain, and are adapted during a warmup phase to obtain an acceptance ratio in the range $[0.2, 0.3]$. Physically, β_d can be seen as favoring changes close to the right edge of the bulk, while Δ favors large norms of $\mathbf{z}_\mu$, more likely to induce macroscopic changes in the largest eigenvalue.

Results of the experiments

The code is available in a public Github repository [Mai21] and the results of the simulations are given in Fig. 8.3 for four different choices of $\rho(t)$. The agreement with Result 8.1 is excellent,

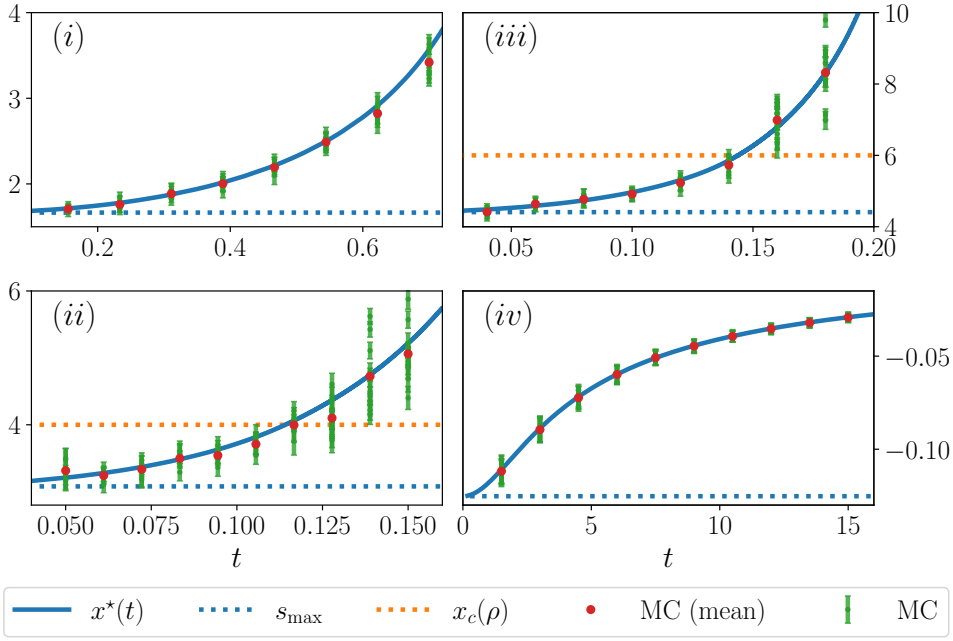


FIGURE 8.3: The function $x^\star(t)$ for $\rho(t) =:$ (i) two peaks $(\delta_1 + \delta_{-1})/2$, (ii) Wigner's semicircle law, (iii) the Marchenko-Pastur law with ratio 1, (iv) the uniform distribution in $[-2, -1]$. In all cases $\alpha = 2$ except for (ii), in which $\alpha = 1$. Solid lines are analytical predictions. The different Monte-Carlo runs ($n = 500$) are shown in green with their respective noise. The mean of the green points is depicted as a red dot.

whether $\mathbf{H}_n$ is negative, positive, or neither. Even though the variability of the results naturally increases with t, we are able to access very large values of $x^\star(t)$, beyond the transition point $x_c(\rho)$ when it exists (cf. cases (ii) and (iii)). For example, in (iii) we are able to sample up to $x^\star(t) \simeq 8$. Comparing with Fig. 8.2, this implies that our simulations reach events which have probability of order $e^{-0.5n} \sim 10^{-109}$ under a naive sampling! We were therefore able to probe Result 8.1 deep into the large deviations regime.

8.4 Derivation of the rate function

In this section, we derive Result 8.1. We will focus on the real case $\beta = 1$, and briefly describe at the end how to generalize our arguments to the complex case.

8.4.1 General idea behind the method

Let $x \geq s_{\max}$, and let us denote $P_n(x)$ the PDF of $\lambda_{\max}(\mathbf{H}_n)$. Our aim is to compute the asymptotics of $n^{-1} \ln P_n(x)$, i.e. the probability of exponentially rare events in which $\lambda_{\max}(\mathbf{H}_n)$ is close to x rather than to its typical value $s_{\max}$.

The main idea of the method is to *tilt* the probability measure of $\mathbf{H}_n$ so that having $\lambda_{\max}(\mathbf{H}_n) \simeq x$ becomes a *typical* event, rather than an exponentially rare one. This new tilted law will be parametrized by a number $\theta \geq 0$: for each θ, the largest eigenvalue will typically be close to a value $x(\theta)$ as n gets large (very similarly to the importance sampling strategy used for Monte-Carlo simulations, cf. eq. (8.4)). Conversely, each $x \geq s_{\max}$ will be associated to a $\theta_x \geq 0$, and a tilting parametrized by θ_x will typically induce the largest eigenvalue to be close to x.

As we will see, these functions $\{x(\theta), \theta_x\}$ contain all the information about the large deviations we want to compute. To put it roughly, studying how much tilt we need to apply in order to

push the largest eigenvalue from $s_{\max}$ to x will give us information on what was the probability that this eigenvalue was close to x in the first place. On a general note, the tilting strategy is a privileged approach to prove many results in large deviations theory, e.g. Cramer's theorem 1.8 [DZ98]. However, its adaptation to the random matrix context, using the spherical integrals we introduced in Section 1.5.3, is very recent and was introduced by the series of mathematical works mentioned above.

8.4.2 Tilting the measure: a first attempt

The tilted law of $\mathbf{H}_n$

We start with a first simple use of the tilting method using spherical integrals[3]. The simplest possible tilting of the measure, inspired by the aforementioned mathematical works, is an exponential tilting. More precisely, we define the tilted distribution of $\mathbf{z}$ as[4]:

$$P_{\theta,\mathbf{e}}(\mathbf{z})\mathrm{d}\mathbf{z} \propto \mathcal{D}\mathbf{z} \exp\Big\{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}\Big\}, \tag{8.6}$$

for a given vector $\mathbf{e} \in \mathbb{S}^{n-1}$ and a parameter $\theta \geq 0$. As we will see, this tilting induces a macroscopic move of the largest eigenvalue. Moreover, by rotation invariance of the distribution of $\mathbf{H}_n$, this move only depends on θ, and not on the specific direction of $\mathbf{e}$.

Let us now detail the distribution of $\mathbf{H}_n$ under the tilted law of eq. (8.6). Computing the normalization factor, we reach:

$$\begin{aligned} P_{\theta,\mathbf{e}}(\mathbf{z}) &= \exp\Big\{\frac{1}{2}\sum_{\mu=1}^m \ln\Big(1-\frac{\theta}{\alpha}d_\mu\Big) + \frac{\theta}{2\alpha}\sum_{\mu=1}^m d_\mu(\mathbf{e}^\intercal\mathbf{z}_\mu)^2 - \frac{1}{2}\sum_{\mu=1}^m \|\mathbf{z}_\mu\|^2 - \frac{nm}{2}\ln 2\pi\Big\}, \\ &= \prod_{\mu=1}^m \exp\Big\{-\frac{1}{2}\mathbf{z}_\mu^\intercal(\mathrm{I}_n - \frac{\theta}{\alpha}d_\mu\mathbf{e}\mathbf{e}^\intercal)\mathbf{z}_\mu - \frac{n}{2}\ln 2\pi + \frac{1}{2}\ln\det\big(\mathrm{I}_n - \frac{\theta}{\alpha}d_\mu\mathbf{e}\mathbf{e}^\intercal\big)\Big\}. \end{aligned} \tag{8.7}$$

The matrix $\mathrm{I}_n - (\theta/\alpha)d_\mu\mathbf{e}\mathbf{e}^\intercal$ is a rank-one modification of the identity, so we easily compute

$$\Big(\mathrm{I}_n - \frac{\theta}{\alpha}d_\mu\mathbf{e}\mathbf{e}^\intercal\Big)^{-1/2} = \mathrm{I}_n + ((1-\theta d_\mu/\alpha)^{-1/2}-1)\mathbf{e}\mathbf{e}^\intercal. \tag{8.8}$$

Changing variables to $\mathbf{z}'_\mu \equiv (\mathrm{I}_n - \theta d_\mu \mathbf{e}\mathbf{e}^\intercal/\alpha)^{1/2}\mathbf{z}_\mu$ in eq. (8.7) and using eq. (8.8) yields that $\mathbf{H}_n$ is distributed under $P_{\theta,\mathbf{e}}$ as:

$$\mathbf{H}_n^{(\mathbf{e},\theta)} \overset{\mathrm{d}}{=} \frac{1}{m}\sum_{\mu=1}^m d_\mu[\mathrm{I}_n + \kappa_\theta(d_\mu)\mathbf{e}\mathbf{e}^\intercal]\mathbf{z}_\mu\mathbf{z}_\mu^\intercal[\mathrm{I}_n + \kappa_\theta(d_\mu)\mathbf{e}\mathbf{e}^\intercal], \tag{8.9}$$

with $\kappa_\theta(t) \equiv (1-\alpha^{-1}\theta t)^{-1/2} - 1$, and in which $\mathbf{z}_\mu$ are again i.i.d. standard Gaussian vectors.

Since $\mathbf{H}_n^{(\mathbf{e},\theta)}$ is a finite-rank change of $\mathbf{H}_n$, its largest eigenvalue can be typically larger than $s_{\max}$. Moreover, we see from the expression of $\mathbf{H}_n^{(\mathbf{e},\theta)}$ that as $\theta \to \alpha/d_{\max}$, some of the coefficients $\kappa_\theta(d_\mu)$ will grow very large: we thus expect that for sufficiently large θ, an outlier eigenvalue will pop out from the right edge of the "bulk", as pictured as a black arrow in Fig. 8.1. We denote $x(\theta) \geq s_{\max}$ the typical value of this outlier eigenvalue (i.e. of $\lambda_{\max}(\mathbf{H}_n)$), as $n \to \infty$. As we mentioned, $x(\theta)$ does not depend on $\mathbf{e}$ by rotation invariance.

[3] The analysis of [BG20, BGH20] suggests a tilting which is function of $\mathbf{A}_n$, with $\mathbf{H}_n = \mathbf{A}_n\mathbf{A}_n^\intercal$. However for arbitrary d_μ, $\mathbf{A}_n$ is not defined so that we use this simpler tilting. We shall later come back to this idea by adapting the method to allow for complex-valued $\mathbf{A}_n$.

[4] Recall that $\mathcal{D}\mathbf{z} \equiv \mathrm{d}\mathbf{z}\ e^{-||\mathbf{z}||^2/2}/(2\pi)^{n/2}$ is the standard Gaussian law.

Relating the tilting and the original problem

Let us now see how to relate the PDF P_n of $\lambda_{\max}(\mathbf{H}_n)$ to this tilted distribution. For any θ, we can write the trivial identity:

$$P_n(x(\theta)) = \int \mathcal{D}\mathbf{z}\ \delta(\lambda_{\max}(\mathbf{H}_n) - x(\theta)) = \int \mathcal{D}\mathbf{z}\ \delta(\lambda_{\max}(\mathbf{H}_n) - x(\theta))\ \frac{\int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ e^{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}}{\int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ e^{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}},$$

$$= \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ \mathcal{D}\mathbf{z}\ \delta(\lambda_{\max}(\mathbf{H}_n) - x(\theta)) \frac{e^{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}}{e^{nJ_n(\mathbf{H}_n,\theta)}}. \tag{8.10}$$

We introduced the *spherical integral*

$$J_n(\mathbf{H}_n, \theta) \equiv \frac{1}{n} \ln \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ e^{\frac{n\theta}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}. \tag{8.11}$$

In eq. (8.10), we almost see the tilted probability distribution of eq. (8.6) appearing. However this is not exactly the case, as the term $e^{nJ_n(\mathbf{H}_n,\theta)}$ depends on $\mathbf{z}$ via $\mathbf{H}_n$. If this term would not depend on $\mathbf{z}$ at leading exponential order in n, the law of $\mathbf{z}$ would be the tilted law of eq. (8.6), and we could remove the δ term in eq. (8.10): indeed, the constraint $\lambda_{\max}(\mathbf{H}_n) \simeq x(\theta)$ would already be satisfied by the very definition of $x(\theta)$!

The spherical integrals

Therefore, we study first $J_n(\mathbf{H}_n, \theta)$. Let us introduce $J_1(\theta, x)$, defined as the limit of $J_n(\mathbf{H}_n, \theta)$, *assuming* $\lambda_{\max}(\mathbf{H}_n) \to x$ as $n \to \infty$ (which we can safely assume because of the constraint in eq. (8.10)). Adopting the language of statistical physics, we call J_1 a *quenched* spherical integral. More precisely, J_1 belongs to a class of high-dimensional integrals known as Harish-Chandra-Itzykson-Zuber (HCIZ) integrals [HC57, IZ80], that we already introduced and studied in Section 1.5.3. In particular, Theorem 1.12 yields:

$$J_1(\theta, x) = \inf_{\gamma > \theta x} \left[\frac{\gamma}{2} - \frac{1}{2}\int \mathrm{d}u\, \sigma(u) \ln(\gamma - \theta u)\right] - \frac{1}{2}. \tag{8.12}$$

Let us come back to eq. (8.10). We have at leading exponential order:

$$P_n(x(\theta)) \simeq \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ \mathcal{D}\mathbf{z}\ \delta(\lambda_{\max}(\mathbf{H}_n) - x(\theta)) \frac{e^{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}}{e^{nJ_1(\theta, x(\theta))}},$$

$$\simeq e^{-nJ_1(\theta,x(\theta))} \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ \mathcal{D}\mathbf{z}\ e^{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}. \tag{8.13}$$

As already argued, we removed the δ constraint in eq. (8.13) by definition of $x(\theta)$: under the tilted law $P_{\theta,\mathbf{e}}$, the largest eigenvalue $\lambda_{\max}(\mathbf{H}_n)$ typically concentrates on $x(\theta)$, so this constraint is superfluous. The expression of eq. (8.13) involves another integral, that we call *annealed* and denote $F_n(\theta)$, borrowing again from the statistical physics jargon:

$$F_n(\theta) \equiv \frac{1}{n} \ln \int \mathcal{D}\mathbf{z} \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ e^{\frac{\theta n}{2}\mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}.$$

Similarly to J_n, we denote by $F_1(\theta)$ the limit of $F_n(\theta)$. If $d_{\max} > 0$, we also impose $\theta < \alpha/d_{\max}$ so that $F_n(\theta)$ is well-defined. We compute it by direct integration on $\mathbf{z}$:

$$F_1(\theta) = -\frac{\alpha}{2} \int \mathrm{d}t \rho(t) \ln(1 - \alpha^{-1}\theta t). \tag{8.14}$$

Combined with eq. (8.13), this implies

$$P_n(x(\theta)) \simeq \exp\{-n[J_1(\theta, x(\theta)) - F_1(\theta)]\}. \tag{8.15}$$

Note that we imposed $\theta < \theta_{\max}$, with

$$\theta_{\max} \equiv \begin{cases} \alpha/d_{\max} & \text{if } d_{\max} > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Conversely, this implies that eq. (8.15) can only be applied for $s_{\max} \leq x < x_{\max} \equiv x(\theta_{\max})$. This creates a possibly important limitation of the tilting we used, if $x_{\max}$ is finite: in this case, the method does not give access to the large deviations for $x \geq x_{\max}$! We will precisely characterize when such a limitation occurs in the following, relating it to the phase transition phenomenon described in Section 8.2, and we will develop a second tilting to circumvent this issue.

Simplifying the rate function

First, let us focus on $x < x_{\max}$ and show that we find Result 8.1. We can rewrite eq. (8.15) as:

$$\frac{1}{n} \ln P_n(x) \simeq -[J_1(\theta_x, x) - F_1(\theta_x)]. \tag{8.16}$$

Recall that θ_x is chosen exactly to be able to remove the delta constraint in eq. (8.10). However, for any $\theta' \geq 0$, we can always write an equivalent to eq. (8.13), keeping the delta constraint:

$$P_n(x) \simeq \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e}\ \mathcal{D}\mathbf{z}\ \delta(\lambda_{\max}(\mathbf{H}_n) - x) \frac{e^{\frac{\theta' n}{2} \mathbf{e}^\intercal \mathbf{H}_n \mathbf{e}}}{e^{n J_1(\theta', x)}}.$$

From here, we can upper bound $P_n(x)$ by discarding the delta constraint in this equation, which gives at leading exponential order:

$$\frac{1}{n} \ln P_n(x) \lessapprox -[J_1(\theta', x) - F_1(\theta')].$$

Combining this with eq. (8.16), we see we can write the rate function $I(x) \simeq n^{-1} \ln P_n(x)$ as:

$$I(x) = \sup_{\theta \in [0, \theta_{\max})} [J_1(\theta, x) - F_1(\theta)]. \tag{8.17}$$

We focus now on simplifying the rate function of eq. (8.17), to obtain Result 8.1. We need to study the behavior of the quenched integral J_1 of eq. (8.12). We recall here known results on $J_1(\theta, x)$, as we already studied the behavior of such a function, and the possible transitions in it, in Section 1.5.3.

Cancelling the derivative with respect to γ in eq. (8.12) yields $\gamma = \gamma^\star \equiv \theta G_\sigma^{-1}(\theta)$. Plugging back this solution in eq. (8.12), and using the Marchenko-Pastur equation (8.1) yields that $J_1(\theta, x) = F_1(\theta)$. We detail this derivation in Appendix E.5.1.

However, note that γ is constrained to be smaller than θx. Therefore, the infimum in eq. (8.12) is reached in $\gamma^\star$ only for $\theta \leq \theta_c(x)$, with $\theta_c(x) \equiv G_\sigma(x)$. At $\theta = \theta_c(x)$, $J_1(\theta, x)$ undergoes a transition, as γ "saturates" at its limit value θx for $\theta \geq \theta_c(x)$. All in all, we reach:

$$J_1(\theta, x) = \begin{cases} F_1(\theta) = -\dfrac{\alpha}{2} \displaystyle\int \mathrm{d}t \rho(t) \ln(1 - \alpha^{-1}\theta t) & \text{if } \theta \leq G_\sigma(x), \\ \dfrac{\theta x - 1 - \ln \theta}{2} - \dfrac{1}{2} \displaystyle\int \mathrm{d}u\ \sigma(u) \ln(x - u) & \text{if } \theta \geq G_\sigma(x). \end{cases} \tag{8.18}$$

Using eq. (8.18) in the result of eq. (8.17), it is then simple algebra to see that the maximum of $J_1(\theta, x) - F_1(\theta)$ is reached in $\theta_x = \overline{G}_\sigma(x)$. Differentiating the resulting expression yields

$$I'(x) = \frac{\overline{G}_\sigma(x) - G_\sigma(x)}{2},$$

which justifies Result 8.1 in this case. The whole computation we just described is detailed in Appendix E.6.1.

Limitations of the tilting

As we mentioned, the tilting method we presented is not capable of predicting the large deviations for $x \geq x_{\max} = x(\theta_{\max})$. As we showed that $\theta_x = \overline{G}_\sigma(x)$, we can separate two cases:

- If $d_{\max} \leq 0$, then $\theta_{\max} = +\infty$ by definition, and therefore $x_{\max} = 0$ since $\lim_{x \uparrow 0} \overline{G}_\sigma(x) = +\infty$. For $x \geq 0$, $\overline{G}_\sigma(x) = +\infty$ and so Result 8.1 is valid (indeed $I(x) = +\infty$ since $\mathbf{H}_n$ is negative). In the end, our tilting allowed to compute the large deviations rate function $I(x)$ for any $x \geq s_{\max}$ in this case.
- If $d_{\max} > 0$, then $\theta_{\max} = \alpha/d_{\max}$. Since $\overline{G}_\sigma(x_c(\rho)) = \alpha/d_{\max}$, this yields that $x_{\max} = x_c(\rho)$, given by eq. (8.3). Therefore, we see that in this case, the condition for the tilting to be able to induce arbitrarily large outliers is $x_c(\rho) = +\infty$, i.e. $G_\rho(d_{\max}) = +\infty$. As we saw, the finiteness of $G_\rho(d_{\max})$ is exactly the existence condition of a phase transition in $I(x)$, which prevents the tilting from capturing all the large deviations.

8.4.3 Beyond the transition: a second tilting

Here, we briefly outline the method we use to go beyond the phase transition when $d_{\max} > 0$, to circumvent the limitation described above. As the method is extremely similar to the one we just described in detail, we will focus primarily on the main steps and quantities, while leaving some details to the reader. We change the tilt of eq. (8.6) to:

$$P_{\theta,\mathbf{e},\mathbf{f}}(\mathbf{z})\,\mathrm{d}\mathbf{z} \propto \mathcal{D}\mathbf{z} \exp\Big\{\frac{\theta n}{\sqrt{m}} \sum_{i,\mu} \sqrt{d_\mu} e_i z_{\mu i} f_\mu\Big\}, \tag{8.19}$$

with $\sum_i e_i^2 = \sum_\mu f_\mu^2 = 1$. When $d_\mu \leq 0$, we define $\sqrt{d_\mu} \equiv i\sqrt{-d_\mu}$ so that the tilt is possibly complex-valued. Eq. (8.19) corresponds to a simple additive shift of $\mathbf{z}_\mu$, and the law of $\mathbf{H}_n$ under the tilt of eq. (8.19) is:

$$\mathbf{H}_n^{(\theta,\mathbf{e},\mathbf{f})} \equiv \frac{1}{m} \sum_{\mu=1}^{m} \Big[d_\mu \mathbf{z}_\mu \mathbf{z}_\mu^\intercal + \frac{\theta^2 m}{\alpha^2} d_\mu^2 f_\mu^2 \mathbf{e}\mathbf{e}^\intercal + \frac{\theta\sqrt{m}}{\alpha} \mathbb{1}_{\{d_\mu \geq 0\}} d_\mu^{3/2} f_\mu (\mathbf{e}\mathbf{z}_\mu^\intercal + \mathbf{z}_\mu \mathbf{e}^\intercal) \Big].$$

Let us give an intuitive view of the reasons why this new tilting manages to induce the largest eigenvalue to be typically close to x, for any $x \geq s_{\max}$:

- When $\theta = 0$ the largest eigenvalue of the unspiked matrix naturally concentrates on $s_{\max}$.
- As $\theta \gg 1$, a spike proportional to θ^2 will push the largest eigenvalue of $\mathbf{H}_n^{(\theta,\mathbf{e},\mathbf{f})}$ to $+\infty$.

By continuously varying θ, we see that the tilt should be able to induce any outlier $x \geq s_{\max}$ in the spectrum. The annealed and quenched "HCIZ" integrals corresponding to this tilting are:

$$F_2(\theta) = \frac{1}{n} \ln \int \mathcal{D}\mathbf{z} \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e} \int_{\|\mathbf{f}\|^2=1} \mathrm{d}\mathbf{f} \, \exp\Big\{ \frac{\theta n}{\sqrt{m}} \sum_{i,\mu} \sqrt{d_\mu} e_i z_{\mu i} f_\mu \Big\},$$
$$J_2(\theta, x) = \frac{1}{n} \ln \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e} \int_{\|\mathbf{f}\|^2=1} \mathrm{d}\mathbf{f} \, \exp\Big\{ \frac{\theta n}{\sqrt{m}} \sum_{i,\mu} \sqrt{d_\mu} e_i z_{\mu i} f_\mu \Big\}.$$

In $J_2(\theta, x)$, we assume that $\lambda_{\max}(\mathbf{H}_n)$ converges a.s. to x as $n \to \infty$. Introducing Lagrange multipliers in the spherical integrals, we find:

$$F_2(\theta) = \frac{\alpha}{2} \inf_{\gamma \geq d_{\max}} \Big[\frac{\gamma \theta^2}{\alpha^2} - \int \mathrm{d}t \rho(t) \ln(\gamma - t) - 1 - \ln \frac{\theta^2}{\alpha^2} \Big].$$

Similarly to our previous analysis of J_1, we show that there is a transition in J_2: for $\theta \leq \theta_c(x)$, $J_2(\theta, x) = F_2(\theta)$, while for $\theta \geq \theta_c(x)$ one reaches:

$$J_2(\theta, x) = \begin{cases} F_2(\theta) & \text{if } \theta \leq \theta_c(x), \\ \dfrac{\alpha - 1}{2} \ln \Big[\dfrac{1 - \alpha + \sqrt{(\alpha-1)^2 + 4x\theta^2}}{2x} \Big] - \dfrac{1+\alpha}{2} - \dfrac{\alpha}{2} \ln \dfrac{\theta^2}{\alpha} & \\ \qquad + \dfrac{1}{2} \sqrt{(\alpha-1)^2 + 4x\theta^2} - \dfrac{1}{2} \displaystyle\int \mathrm{d}\lambda \, \sigma(\lambda) \ln(x - \lambda) & \text{if } \theta \geq \theta_c(x), \end{cases}$$

with $\theta_c(x) \equiv \sqrt{x G_\sigma(x)^2 + (\alpha - 1) G_\sigma(x)}$. The details of the derivations of F_2 and J_2 are given in Appendix E.5.2. Importantly, the very existence of the transition in $J_2(\theta, x)$ relies on the positivity of x, so that this tilting fails for negative matrices. This notably implies that the first tilt of eq. (8.6) is still crucial to handle the case $d_{\max} \leq 0$.

We deduce from the tilting method, in the exact same way as in the first tilting, that

$$P_n(x) \simeq \exp\{ -n \sup_{\theta \geq 0} [J_2(\theta, x) - F_2(\theta)] \}.$$

Using eq. (8.1) and the explicit expressions of F_2 and J_2 we derived, one shows that for all $x \geq s_{\max}$ the supremum is attained in $\theta_x \equiv [x \overline{G}_\sigma(x)^2 + (\alpha - 1) \overline{G}_\sigma(x)]^{1/2}$. We compute then $I'(x) = [\overline{G}_\sigma(x) - G_\sigma(x)]/2$, which, together with $I(s_{\max}) = 0$, implies Result 8.1. These algebraic calculations are detailed in Appendix E.6.2. This ends the derivation of Result 8.1 in all cases.

8.4.4 Going further: the complex case and the left tail of the large deviations

A remark on the complex case

We give an intuitive remark on how the factor $\beta = 2$ in the complex case arises in Result 8.1. As we showed above, the method allows to write the large deviations rate function in the form $I(x) = \sup_\theta [J(\theta, x) - F(\theta)]$, with F and J annealed and quenched spherical integrals. This result straightforwardly transfers to the complex setting, however the integrals F and J are now defined over unit vectors on the *complex* unit sphere, i.e. they satisfy $\sum_i |e_i|^2 = 1$. It is known that the asymptotic behavior of real and complex spherical integrals only differ by a factor 2 (i.e. the complex integral is twice the real one), a property known as "Zuber's 1/2-rule"[ZJZ03]: this explains the origin of the β factor in Result 8.1.

The left tail of the large deviations

Importantly, we did not consider large deviations at the left of $s_{\max}$. Such an event requires moving the whole bulk of eigenvalues, i.e. a number $\mathcal{O}(n)$ of eigenvalues, an event which has probability in the scale $\exp\{-n^2\}$ [DM06, VMB07, MV09]. Whether the method applied here could be extended to study this left tail is an interesting open question. As we saw, the core of the method is to create a tilt of the measure such that the largest eigenvalue is shifted in a controllable manner: in this case, the tilting would need to induce a shift of the whole spectrum. The perhaps most natural extension of the tilting of eq. (8.6) to this setting would be to consider an extensive-rank change in the covariance of the $\mathbf{z}_\mu$:

$$\mathcal{D}\mathbf{z} \to \mathcal{D}\mathbf{z}\; e^{\frac{n}{2}\mathrm{Tr}[\mathbf{M}_n \mathbf{O}\mathbf{H}_n\mathbf{O}^\intercal]},$$

with $\mathbf{O}$ an orthogonal matrix and $\mathbf{M}_n$ an arbitrary matrix (with extensive rank) that will parametrize the tilting, similarly to the parameter θ in the calculation we performed. Provided the mechanisms of the method we presented transfer to this case, this would give the large deviations function in terms of involved *extensive-rank* "HCIZ" spherical integrals. The study of these extensive-rank HCIZ integrals in the high-dimensional limit was conducted in [Mat94], and rigorously proven in [GZ02], and we stated their main result in Theorem 1.15. The resulting formulas are very tedious as they involve hydrodynamical PDEs, however in a very recent analysis [BGH20] the authors managed to leverage this approach to prove large deviations in several contexts, e.g. for the empirical measure of the sum of two freely independent random matrices. This indicates that these extensive-rank spherical integrals might indeed be the most natural path to analyse the left tail of the large deviations, a line of work that is however beyond the scope of this chapter.

Discussion on Chapter 8

In this chapter we presented a generic technique to derive the right tail of the large deviations of the largest eigenvalue of the large class of *generalized sample covariance matrices*. By symmetry, this also transfers to the left tail of the large deviations of the smallest eigenvalue. Our main result 8.1 significantly improves over the seminal works of [DM06, MV09] for sample covariance matrices with identity covariance, and leverages a recent technique developed in a series of mathematical works [BG20, GH20, Hus20, BGH20, AGH21, McK21b]. Thanks to the relative simplicity of our main result, we will further investigate its consequences in particular for PCA on real-world datasets. We also proposed importance sampling simulations, that allow to probe our result and its consequences, e.g. the existence of a phase transition in the rate function depending on the spectral density of a diagonal matrix involved in the definition of our model.

The Hessian matrix of inference landscapes – Recall that we motivated Result 8.1 in particular by its application to the Kac-Rice formalism developed in Chapter 7. Our analytical large deviations computation therefore paves the way toward a direct understanding of the topology of the local minima in the landscape of complex inference models, since $\mathbf{H}_n$ is directly related to the Hessian matrix of the models studied in Chapter 7! We are currently investigating the derivation of a Kac-Rice formula similar to Theorem 7.5 but restricted to local minima, which will contain the rate function of Result 8.1.

Universality of the large deviations – Very interestingly the tilting method used in this chapter does not fundamentally rely on the Gaussianity of the variables $z_{\mu i}$, and several works have used this technique to investigate universality properties of the large deviations of the extreme eigenvalues, e.g. for Wigner matrices [GH20, Hus20, AGH21]. This hints towards a possible universality of the large deviations described in Result 8.1 as long as the variables $z_{\mu i}$

are independent, centered, have unit variance, and satisfy some properties, e.g. sub-Gaussianity or if they are Rademacher/Ising variables. Proving or refuting this intuition in detail would be an interesting follow-up to this chapter.

Other possible generalizations – An interesting result, which should be accessible, is to generalize the large deviations of Result 8.1 to the k-th eigenvalue (not necessarily the smallest), with fixed k as $n \to \infty$. In terms of the spherical integrals, which are at the heart of the tilting techniques (cf. Section 8.4), this requires results on the asymptotics of rank-k HCIZ spherical integrals, which are well-known, see Section 1.5.3. This venue has recently been investigated for Wigner and Wishart matrices [GH21], and generalizing to the present context should follow on the same lines. In the Kac-Rice formalism of Chapter 7, this would then allow to count the critical points of any finite index (i.e. saddle points with a finite number of descending directions). As a final note, the generalizations mentioned in the conclusion to Chapter 7 would naturally create new random matrix challenges, e.g. understanding the large deviations of the extremal eigenvalues of the Hessian of the loss for a neural network with multiple layers and nodes, a random matrix problem that, to the best of our knowledge, is still open at the moment.

Afterword

"Look at me still talking when there's science to do"

GLaDOS, Portal (2007).

This dissertation aimed at leveraging diverse tools of statistical physics, statistics, information theory and probability theory to tackle different questions on the fundamental limits of estimation problems. In Part I we revisited high-temperature expansions, historically one of the first theoretical methods developed to deal with disordered systems, to unify and justify different algorithmic approaches for inference problems with rotationally-invariant data distributions. We also used it to provide first hints at an exact solution to the extensive-rank matrix factorization problem, which puts into the light the limitations of current theories, as we will further discuss below. Part II places itself in the general line of work reviewed in [ZK16], and which combines the heuristic replica method and message-passing algorithms to assess the optimal performances in a wide class of high-dimensional estimation problems. We complete this approach with probabilistic methods, notably an adaptive interpolation developed in [BM19a, BM19b], which allows to put the replica predictions on rigorous grounds in quite general settings. This general venue is applied to two-layers neural networks (Chapter 4), to study the influence of data structure on the optimal performances (Chapter 5), and to analyze the phase retrieval problem with generic right-rotationally invariant sensing matrices (Chapter 6). Finally, Part III considers a topological point of view on the general question of high-dimensional optimization, and provides an important step towards a mathematical characterization of the ruggedness of the landscapes of empirical risk minimization for simple estimation models, with important connections to random matrix theory.

Towards a global theory of learning? – Nonetheless, this dissertation touches upon important limitations of the statistical physics approach to high-dimensional estimation. Crucially, as the rest of "classical" approaches to machine learning, it fails to provide a clear theory of modern machine learning procedures based on deep neural networks, namely to answer the following problems that remain open:

- Why do models with very large number of parameters (such as deep nets), often larger than the number of data, not overfit? How do they manage to generalize well?
- Why do the optimization algorithms used in such models not get stuck on poor local minima?

These questions defy our statistical intuition, and as emphasized in [Zde20], answering them can not be done by overlooking any of the three following key ingredients: the architecture of the problem, the structure of the data, and the learning algorithms. In this thesis we were confronted with each of these features, and an important long-term goal is to unify what statistical physics taught us on each of them to obtain answers.

We can illustrate this strategy on the analysis of data structure: indeed, the consideration of i.i.d. data samples has historically been one of the important limitations of the statistical physics approach, but recent progress has been made on going beyond this restriction, either by rotationally-invariant models or generative models as shown in Parts I and II, and in particular

generative models of data are better and better understood using statistical physics techniques [GRM+20, GMKZ20]. This is also the case of optimization algorithms: while physicists have described the Langevin dynamics in continuous disordered models very precisely [CK93], this description is not adapted to study actual optimization algorithms (e.g. stochastic gradient descent) in machine learning procedures[1]. However, recent works in the same community strive to create an analytical description of realistic optimization algorithms using statistical physics approaches (let us mention e.g. [MKUZ20]), and combining these diverse ideas is an exciting prospect.

The extensive-rank challenge – Let us conclude our discussion by what we believe is a crucial challenge ahead for the statistical physics community interested in inference and learning: the analysis of *extensive-rank* problems, to which we gave a brief introduction in Chapter 3. Such models are especially instrumental in building a theory of actual deep neural networks with both many layers and many nodes in each layer, and the tremendous attention received by simple learning regimes in infinitely-wide networks [JGH18, GMMM21] demonstrates the importance of these models to the community. However, the tools that have been developed for fifty years by theoretical physicists and statisticians, such as the replica theory, are often inadequate to tackle extensive-rank problems as we explained in Chapter 3, and new ideas must be developed to expand them or to develop new ones. Finally, even though we focused in the aforementioned chapter on the matrix factorization problem with i.i.d. Gaussian weights, this model is an instrumental building block to move on to structured data and the design of efficient algorithms, much like the perceptron with Gaussian weights in finite-rank problems. While this very exciting research avenue is surely a very long-term goal, it is also a thrilling opportunity to rethink many of our intuitions and to gather again various scientific communities in physics, computer science, statistics, and probability.

[1] This illustrates the arguably greatest peril of our interdisciplinary approach: to perform careless transposition of results from one scientific field to the other.

Bibliography

PhD publications

[ALM+20] Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. The spiked matrix model with generative priors. *IEEE Transactions on Information Theory*, 2020.

[AMB+19] Benjamin Aubin, Antoine Maillard, Jean Barbier, Florent Krzakala, Nicolas Macris, and Lenka Zdeborová. The committee machine: computational to statistical gaps in learning a two-layers neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124023, 2019.

[BMMK18] Jean Barbier, Nicolas Macris, Antoine Maillard, and Florent Krzakala. The mutual information in random linear estimation beyond iid matrices. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1390–1394. IEEE, 2018.

[Mai19] Antoine Maillard. An introduction to the Kac-Rice formula. Notes of a short course given at KITP, 2019.

[Mai21] Antoine Maillard. Large deviations of extreme eigenvalues of generalized sample covariance matrices. *EPL (Europhysics Letters)*, 133(2):20005, 2021.

[MBAB20] Antoine Maillard, Gérard Ben Arous, and Giulio Biroli. Landscape complexity for the empirical risk of generalized linear models. In *Mathematical and Scientific Machine Learning*, pages 287–327. PMLR, 2020.

[MFC+19] Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. High-temperature expansions and message-passing algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(11):113301, 2019.

[MFK+21] Antoine Maillard, Laura Foini, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Towards exact solution of extensive-rank matrix factorization. *In preparation*, 2021.

[MKLZ21] Antoine Maillard, Florent Krzakala, Yue M. Lu, and Lenka Zdeborová. Construction of optimal spectral methods in phase retrieval. In *Mathematical and Scientific Machine Learning*. PMLR, 2021.

[MLKZ20] Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Phase retrieval in high dimensions: Statistical and computational phase transitions. *Advances in Neural Information Processing Systems*, 33, 2020.

Numerical codes of PhD publications

[ALM+19] Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. Demonstration codes - The spiked matrix model with generative priors. https://github.com/benjaminaubin/StructuredPrior_demo, 2019.

[AMB+18] Benjamin Aubin, Antoine Maillard, Jean Barbier, Florent Krzakala, Nicolas Macris, and Lenka Zdeborová. AMP implementation of the committee machine. https://github.com/benjaminaubin/TheCommitteeMachine, 2018.

[Mai21] Antoine Maillard. Demonstration codes - Large deviations of extreme eigenvalues of generalized sample covariance matrices. https://github.com/AnMaillard/LD_lmax_sample_covariance, 2021.

[MKLZ20] Antoine Maillard, Florent Krzakala, Yue M. Lu, and Lenka Zdeborová. Demonstration codes and notebooks. https://github.com/AnMaillard/Optimal_Spectral_Methods_PR, 2020.

[MLKZ20] Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Demonstration codes and notebooks. https://github.com/sphinxteam/PhaseRetrieval_demo, 2020.

Other references

[AAKZ20] Alia Abbara, Benjamin Aubin, Florent Krzakala, and Lenka Zdeborová. Rademacher complexity and spin glasses: a link between the replica and statistical theories of learning. In *Mathematical and Scientific Machine Learning*, pages 27–54. PMLR, 2020.

[ABA13] Antonio Auffinger and Gérard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *Annals of Probability*, 41(6):4214–4247, 2013.

[ABAČ13] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013.

[Abb20] Alia Abbara. *Statistical mechanics of learning with correlated patterns*. PhD thesis, École Normale Supérieure, 2020.

[AFP16] Ada Altieri, Silvio Franz, and Giorgio Parisi. The jamming transition in high dimension: an analytical study of the TAP equations and the effective thermodynamic potential. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(9):093301, 2016.

[AGH21] Fanny Augeri, Alice Guionnet, and Jonathan Husson. Large deviations for the largest eigenvalue of sub-Gaussian matrices. *Communications in Mathematical Physics*, pages 1–54, 2021.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118. Cambridge university press, 2010.

[ALB+20] Benjamin Aubin, Bruno Loureiro, Antoine Baker, Florent Krzakala, and Lenka Zdeborová. Exact asymptotics for phase retrieval and compressed sensing with

random generative priors. In *Mathematical and Scientific Machine Learning*, pages 55–73. PMLR, 2020.

[Alt18] Ada Altieri. Higher-order corrections to the effective potential close to the jamming transition in the perceptron model. *Physical Review E*, 97(1):012103, 2018.

[AM20] Ahmed El Alaoui and Andrea Montanari. Algorithmic thresholds in mean field spin glasses. *arXiv preprint arXiv:2009.11481*, 2020.

[AMS20] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *arXiv preprint arXiv:2001.00904*, 2020.

[And89] Philip W Anderson. Spin glass vi: Spin glass as cornucopia. *Physics Today*, 42(9):9, 1989.

[ARK+19] Ahmed El Alaoui, Aaditya Ramdas, Florent Krzakala, Lenka Zdeborová, and Michael I Jordan. Decoding from pooled data: Sharp information-theoretic bounds. *SIAM Journal on Mathematics of Data Science*, 1(1):161–188, 2019.

[AT09] R.J. Adler and J.E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer New York, 2009.

[AW09a] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. *The Annals of Statistics*, pages 2877–2921, 2009.

[AW09b] Jean-Marc Azaïs and Mario Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons, 2009.

[AW10] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[BABM21a] Gérard Ben Arous, Paul Bourgade, and Benjamin McKenna. Exponential growth of random determinants beyond invariance. *arXiv preprint arXiv:2105.05000*, 2021.

[BABM21b] Gérard Ben Arous, Paul Bourgade, and Benjamin McKenna. Landscape complexity beyond invariance and the elastic manifold. *arXiv preprint arXiv:2105.05051*, 2021.

[BADG06] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-Kurchan equations for dynamics of spin-glasses. *Probability theory and related fields*, 136(4):619–660, 2006.

[BAG97] Gérard Ben Arous and Alice Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probability theory and related fields*, 108(4):517–542, 1997.

[BAGJ20] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Algorithmic thresholds for tensor PCA. *Annals of Probability*, 48(4):2052–2087, 2020.

[BAGJ21] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Online stochastic gradient descent on non-convex losses from high-dimensional inference. *Journal of Machine Learning Research*, 22(106):1–51, 2021.

[Bah96] Safi R Bahcall. Random matrix model for superconductors in a magnetic field. *Physical review letters*, 77(26):5276, 1996.

[BAKZ20] Antoine Baker, Benjamin Aubin, Florent Krzakala, and Lenka Zdeborová. TRAMP: Compositional inference with TRee Approximate Message Passing. *arXiv preprint arXiv:2004.01571*, 2020.

[BAMMN19] Gérard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019.

[Bar19] Jean Barbier. Overlap matrix concentration in optimal Bayesian inference. *Information and Inference: A Journal of the IMA*, 2019.

[BASZ20] Gérard Ben Arous, Eliran Subag, and Ofer Zeitouni. Geometry and temperature chaos in mixed spherical spin glasses at low temperature: the perturbative regime. *Communications on Pure and Applied Mathematics*, 73(8):1732–1828, 2020.

[Bay63] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 1763.

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[BBBZ07] Carlo Baldassi, Alfredo Braunstein, Nicolas Brunel, and Riccardo Zecchina. Efficient supervised learning in networks with binary synapses. *Proceedings of the National Academy of Sciences*, 104(26):11079–11084, 2007.

[BBVS00] Naama Brenner, William Bialek, and Rob de Ruyter Van Steveninck. Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702, 2000.

[BCMN14] Afonso S Bandeira, Jameson Cahill, Dustin G Mixon, and Aaron A Nelson. Saving phase: Injectivity and stability for phase retrieval. *Applied and Computational Harmonic Analysis*, 37(1):106–125, 2014.

[BCRT20] Giulio Biroli, Chiara Cammarota, and Federico Ricci-Tersenghi. How to iron out rough landscapes and get optimal performances: averaged gradient descent and its application to tensor PCA. *Journal of Physics A: Mathematical and Theoretical*, 53(17):174003, 2020.

[BD07] Alan J Bray and David S Dean. Statistics of critical points of Gaussian fields on large-dimensional spaces. *Physical review letters*, 98(15):150201, 2007.

[BDM+16] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems*, pages 424–432, 2016.

[BG11] Florent Benaych-Georges. Rectangular R-transform as the limit of rectangular spherical integrals. *Journal of Theoretical Probability*, 24(4):969, 2011.

[BG20] Giulio Biroli and Alice Guionnet. Large deviations for the largest eigenvalues and eigenvectors of spiked Gaussian random matrices. *Electronic Communications in Probability*, 25, 2020.

[BGH20] Serban Belinschi, Alice Guionnet, and Jiaoyang Huang. Large deviation principles via spherical integrals. *arXiv preprint arXiv:2004.07117*, 2020.

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[BGS84] Oriol Bohigas, Marie-Joya Giannoni, and Charles Schmit. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Physical Review Letters*, 52(1):1, 1984.

[BHL+02] Wolfgang Barthel, Alexander K Hartmann, Michele Leone, Federico Ricci-Tersenghi, Martin Weigt, and Riccardo Zecchina. Hiding solutions in random satisfiability problems: A statistical mechanics approach. *Physical review letters*, 88(18):188701, 2002.

[BJPD17] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 537–546. JMLR. org, 2017.

[BJSG+18] Marco Baity-Jesi, Levent Sagun, Mario Geiger, Stefano Spigler, Gérard Ben Arous, Chiara Cammarota, Yann LeCun, Matthieu Wyart, and Giulio Biroli. Comparing dynamics: Deep neural networks versus glassy systems. In *International Conference on Machine Learning*, pages 314–323. PMLR, 2018.

[BK17] Jean Barbier and Florent Krzakala. Approximate message-passing decoder and capacity achieving sparse superposition codes. *IEEE Transactions on Information Theory*, 63(8):4894–4927, 2017.

[BKM+19] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, page 201802705, 2019.

[BM80] Alan J Bray and MA Moore. Broken replica symmetry and metastable states in spin glasses. *Journal of Physics C: Solid State Physics*, 13(31):L907, 1980.

[BM02] Peter L Bartlett and Shahar Mendelson. Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

[BM11] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[BM19a] Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference. *Probability theory and related fields*, 174(3):1133–1185, 2019.

[BM19b] Jean Barbier and Nicolas Macris. The adaptive interpolation method for proving replica formulas. applications to the Curie–Weiss and Wigner spike models. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294002, 2019.

[BMDK20] Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Transactions on Information Theory*, 66(7):4270–4303, 2020.

[BMN20] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1):33–79, 2020.

[BNS16] Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Global optimality of local search for low rank matrix recovery. In *Advances in Neural Information Processing Systems*, pages 3873–3881, 2016.

[Bol98] Ludwig Boltzmann. *Lectures on gas theory*. J.A. Barth, 1896-1898.

[Bol14] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

[Bor14] Emile Borel. *Introduction géométrique à quelques théories physiques*. Paris, 1914.

[Bor19] Charles Bordenave. Lecture notes on random matrix theory, 2019.

[Bot03] Léon Bottou. Stochastic learning. In *Summer School on Machine Learning*, pages 146–168. Springer, 2003.

[BP00] Jean-Philippe Bouchaud and Marc Potters. *Theory of financial risks*. Cambridge University Press, 2000.

[BP19] Lucas Benigni and Sandrine Péché. Eigenvalue distribution of nonlinear models of random matrices. *arXiv preprint arXiv:1904.03090*, 2019.

[BPW18] Afonso S Bandeira, Amelia Perry, and Alexander S Wein. Notes on computational-to-statistical gaps: Predictions using statistical physics. *Portugaliae Mathematica*, 75(2):159–186, 2018.

[BR13] Quentin Berthet and Philippe Rigollet. Computational lower bounds for sparse PCA. *arXiv preprint arXiv:1304.0828*, 2013.

[BR20] Jean Barbier and Galen Reeves. Information-theoretic limits of a multiview low-rank symmetric spiked matrix model. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2771–2776. IEEE, 2020.

[BS10] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

[Buc13] James Bucklew. *Introduction to rare event simulation*. Springer Science & Business Media, 2013.

[CC05] Tommaso Castellani and Andrea Cavagna. Spin-glass theory for pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05012, 2005.

[CCS+19] Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer Chayes, Levent Sagun, and Riccardo Zecchina. Entropy-SGD: Biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124018, 2019.

[CGG99] Andrea Cavagna, Juan P Garrahan, and Irene Giardina. Quenched complexity of the mean-field p-spin spherical model with external magnetic field. *Journal of Physics A: Mathematical and General*, 32(5):711, 1999.

[CH67] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[CK93] Leticia F Cugliandolo and Jorge Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173, 1993.

[CLS15a] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, 2015.

[CLS15b] Emmanuel J Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[CMW20] Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. In *Conference on Learning Theory*, pages 1078–1141. PMLR, 2020.

[CMZ19] Robert Coquereaux, Colin Mcswiggen, and Jean-Bernard Zuber. On Horn's problem and its volume function. *Communications in Mathematical Physics*, pages 1–31, 2019.

[ÇO19] Burak Çakmak and Manfred Opper. Convergent dynamics for solving the TAP equations of ising models with arbitrary rotation invariant coupling matrices. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1297–1301. IEEE, 2019.

[COFW16] Burak Cakmak, Manfred Opper, Bernard Henri Fleury, and Ole Winther. Self-averaging Expectation Propagation. In *Advances in Approximate Bayesian Inference: NIPS 2016 Workshop*, 2016.

[CPS21] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy II. *Communications in Mathematical Physics*, 381(1):257–291, 2021.

[Cra38] Harald Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actual. Sci. Ind.*, 736:5–23, 1938.

[CRT06] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[CS92] Andrea Crisanti and H-J Sommers. The spherical p-spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, 1992.

[CS95] A Crisanti and H-J Sommers. Thouless-Anderson-Palmer approach to the spherical p-spin spin glass model. *Journal de Physique I*, 5(7):805–813, 1995.

[CŚ06] Benoît Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

[CŚ07] Benoît Collins and Piotr Śniady. New scaling of Itzykson-Zuber integrals. *Annales de l'IHP Probabilités et statistiques*, 43(2):139–146, 2007.

[CT06] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[DAT78] JRL De Almeida and David J Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983, 1978.

[DB20] Rishabh Dudeja and Milad Bakhshizadeh. Universality of linearized message passing for phase retrieval with structured sensing matrices. *arXiv preprint arXiv:2008.10503*, 2020.

[DBMM20] Rishabh Dudeja, Milad Bakhshizadeh, Junjie Ma, and Arian Maleki. Analysis of spectral methods for phase retrieval with random orthogonal matrices. *IEEE Transactions on Information Theory*, 2020.

[DF07] Qian Du and James E Fowler. Hyperspectral image compression using jpeg2000 and principal component analysis. *IEEE Geoscience and Remote sensing letters*, 4(2):201–205, 2007.

[DJM13] David L Donoho, Iain Johnstone, and Andrea Montanari. Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE transactions on information theory*, 59(6):3396–3433, 2013.

[DLM+15] Angélique Drémeau, Antoine Liutkus, David Martina, Ori Katz, Christophe Schülke, Florent Krzakala, Sylvain Gigan, and Laurent Daudet. Reference-less measurement of the transmission matrix of a highly scattering material using a dmd and phase retrieval techniques. *Optics express*, 23(9):11898–11911, 2015.

[DLT+18] Simon Du, Jason Lee, Yuandong Tian, Aarti Singh, and Barnabas Poczos. Gradient descent learns one-hidden-layer CNN: Don't be afraid of spurious local minima. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1339–1348. PMLR, 2018.

[DM06] David S Dean and Satya N Majumdar. Large deviations of extreme eigenvalues of random matrices. *Physical review letters*, 97(16):160201, 2006.

[DM07] Tobin A Driscoll and Kara L Maki. Searching for rare growth factors using multicanonical monte carlo methods. *SIAM review*, 49(4):673–692, 2007.

[DM14a] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse PCA. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.

[DM14b] Yash Deshpande and Andrea Montanari. Sparse PCA via covariance thresholding. In *Advances in Neural Information Processing Systems*, pages 334–342, 2014.

[DM15] Yash Deshpande and Andrea Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, 2015.

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[DMM20] Rishabh Dudeja, Junjie Ma, and Arian Maleki. Information theoretic limits for phase retrieval with subsampled Haar sensing matrices. *IEEE Transactions on Information Theory*, 2020.

[Don06] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[DS67] Nelson Dunford and Jacob T Schwartz. *Linear operators. 2. Spectral theory: self adjoint operators in Hilbert Space*. Interscience Publ., 1967.

[DY20] Xiucai Ding and Fan Yang. Tracy-Widom distribution for the edge eigenvalues of gram type random matrices. *arXiv preprint arXiv:2008.04166*, 2020.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Applications of mathematics. Springer, 1998.

[EA75] Samuel Frederick Edwards and Phil W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[EAK18] Ahmed El Alaoui and Florent Krzakala. Estimation in the spiked Wigner model: A short proof of the replica formula. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1874–1878. IEEE, 2018.

[EARK+18] Ahmed El Alaoui, Aaditya Ramdas, Florent Krzakala, Lenka Zdeborová, and Michael I Jordan. Decoding from pooled data: Phase transitions of message passing. *IEEE Transactions on Information Theory*, 65(1):572–585, 2018.

[EJ76] Sam F Edwards and Raymund C Jones. The eigenvalue spectrum of a large symmetric random matrix. *Journal of Physics A: Mathematical and General*, 9(10):1595, 1976.

[EMH04] Andreas Engel, Rémi Monasson, and Alexander K Hartmann. On large deviation properties of Erdös–Rényi random graphs. *Journal of Statistical Physics*, 117(3-4):387–426, 2004.

[ESY09] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Annals of Probability*, 37(3):815–852, 2009.

[Eul41] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, pages 128–140, 1741.

[EVdB01] Andreas Engel and Christian Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, 2001.

[Far14] Jacques Faraut. Logarithmic potential theory, orthogonal polynomials, and random matrices. *Modern methods in multivariate statistics, Lecture Notes of CIMPA-FECYT-UNESCO-ANR. Hermann*, 2014.

[FB17] C Daniel Freeman and Joan Bruna. Topology and geometry of half-rectified network optimization. In *5th International Conference on Learning Representations, ICLR 2017*, 2017.

[Fed59] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

[Fie82] James R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

[FLD20a] Yan V Fyodorov and Pierre Le Doussal. Manifolds in a high-dimensional random landscape: Complexity of stationary points and depinning. *Physical Review E*, 101(2):020101, 2020.

[FLD20b] Yan V Fyodorov and Pierre Le Doussal. Manifolds pinned by a high-dimensional random landscape: Hessian at the global energy minimum. *Journal of Statistical Physics*, pages 1–40, 2020.

[FMM21] Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *The Annals of Probability*, 49(1):1–45, 2021.

[FN12] Yan V Fyodorov and Celine Nadal. Critical behavior of the number of minima of a random landscape at the glass transition point and the Tracy-Widom distribution. *Physical review letters*, 109(16):167203, 2012.

[FP16] Silvio Franz and Giorgio Parisi. The simplest model of jamming. *Journal of Physics A: Mathematical and Theoretical*, 49(14):145001, 2016.

[FSW07] Yan V Fyodorov, H-J Sommers, and Ian Williams. Density of stationary points in a high dimensional random energy landscape and the onset of glassy behavior. *JETP Letters*, 85(5):261–266, 2007.

[FT20] YV Fyodorov and R Tublin. Counting stationary points of the loss function in the simplest constrained least-square optimization. *Acta Physica Polonica B*, 51(7):1663, 2020.

[Fuk13] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Elsevier, 2013.

[FW07] Yan V Fyodorov and Ian Williams. Replica symmetry breaking condition exposed by random matrix calculation of landscape complexity. *Journal of Statistical Physics*, 129(5-6):1081–1116, 2007.

[Fyo04] Yan V Fyodorov. Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices. *Physical review letters*, 92(24):240601, 2004.

[Gab20] Marylou Gabrié. Mean-field inference methods for neural networks. *Journal of Physics A: Mathematical and Theoretical*, 53(22):223002, 2020.

[GAK20a] Cédric Gerbelot, Alia Abbara, and Florent Krzakala. Asymptotic errors for high-dimensional convex penalized linear regression beyond Gaussian matrices. In *Conference on Learning Theory*, pages 1682–1713. PMLR, 2020.

[GAK20b] Cedric Gerbelot, Alia Abbara, and Florent Krzakala. Asymptotic errors for teacher-student convex generalized linear models (or: How to prove Kabashima's replica formula). *arXiv preprint arXiv:2006.06581*, 2020.

[GAS+20] Sebastian Goldt, Madhu S Advani, Andrew M Saxe, Florent Krzakala, and Lenka Zdeborová. Dynamics of stochastic gradient descent for two-layer neural networks in the teacher–student setup. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(12):124010, 2020.

[GD89] Elizabeth Gardner and Bernard Derrida. Three unfinished works on the optimal storage capacity of networks. *Journal of Physics A: Mathematical and General*, 22(12):1983, 1989.

[GEG+05] Paul Goldbart, Samuel Frederick Edwards, Paul M Goldbart, Nigel Goldenfeld, and David Sherrington. *Stealing the gold: a celebration of the pioneering physics of Sam Edwards*. Oxford University Press on Demand, 2005.

[GH20] Alice Guionnet and Jonathan Husson. Large deviations for the largest eigenvalue of Rademacher matrices. *Annals of Probability*, 48(3):1436–1465, 2020.

[GH21] Alice Guionnet and Jonathan Husson. Asymptotics of $k-$dimensional spherical integrals and applications. *arXiv preprint arXiv:2101.01983*, 2021.

[Gib02] J Willard Gibbs. *Elementary principles in statistical mechanics*. Charles Scribner's Sons, 1902.

[GJZ17] Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, page 1233–1242, 2017.

[GLK+20] Federica Gerace, Bruno Loureiro, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Generalisation error in learning with random features and the hidden manifold model. In *International Conference on Machine Learning*, pages 3452–3462. PMLR, 2020.

[GLM16] Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*, pages 2973–2981, 2016.

[GM05] Alice Guionnet and Mylène Maida. A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of functional analysis*, 222(2):435–490, 2005.

[GM17] Rong Ge and Tengyu Ma. On the optimization landscape of tensor decompositions. In *Advances in Neural Information Processing Systems*, pages 3656–3666, 2017.

[GMKZ20] Sebastian Goldt, Marc Mézard, Florent Krzakala, and Lenka Zdeborová. Modeling the influence of data structure on learning in neural networks: The hidden manifold model. *Physical Review X*, 10(4):041044, 2020.

[GML+19] Marylou Gabrié, Andre Manoel, Clément Luneau, Jean Barbier, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. Entropy and mutual information in models of deep neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124014, 2019.

[GMMM21] Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Linearized two-layers neural networks in high dimension. *The Annals of Statistics*, 49(2):1029–1054, 2021.

[GPAM+14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[GRM+20] Sebastian Goldt, Galen Reeves, Marc Mézard, Florent Krzakala, and Lenka Zdeborová. The Gaussian equivalence of generative models for learning with two-layer neural networks. *arXiv preprint arXiv:2006.14709*, 2020.

[GS18] Tom Goldstein and Christoph Studer. Phasemax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory*, 64(4):2675–2689, 2018.

[GSV05] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE transactions on information theory*, 51(4):1261–1282, 2005.

[Gue03] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

[GY91] Antoine Georges and Jonathan S Yedidia. How to expand around mean-field theory using high-temperature expansions. *Journal of Physics A: Mathematical and General*, 24(9):2173, 1991.

[Gyö90] Géza Györgyi. First-order transition to perfect generalization in a neural network with binary synapses. *Physical Review A*, 41(12):7097, 1990.

[GZ02] Alice Guionnet and Ofer Zeitouni. Large deviations asymptotics for spherical integrals. *Journal of functional analysis*, 188(2):461–515, 2002.

[GZ17] David Gamarnik and Ilias Zadik. High dimensional linear regression with binary coefficients: Mean squared error and a phase transition. In *Conference on Learning Theory (COLT)*, 2017.

[Har82] P. Hartman. *Ordinary Differential Equations: Second Edition.* Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1982.

[HC57] Harish-Chandra. Differential operators on a semisimple Lie algebra. *American Journal of Mathematics*, 79:87–120, 1957.

[HLDM+18] Alexander K Hartmann, Pierre Le Doussal, Satya N Majumdar, Alberto Rosso, and Gregory Schehr. High-precision simulation of the height distribution for the KPZ equation. *EPL (Europhysics Letters)*, 121(6):67004, 2018.

[HLV18] Paul Hand, Oscar Leong, and Vlad Voroninski. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pages 9136–9146, 2018.

[Hop82] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[HP98] Fumio Hiai and Dénes Petz. Eigenvalue density of the Wishart matrix and large deviations. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 1(04):633–646, 1998.

[Hus20] Jonathan Husson. Large deviations for the largest eigenvalue of matrices with variance profiles. *arXiv preprint arXiv:2002.01010*, 2020.

[HV18] Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. In *Conference On Learning Theory*, pages 970–978, 2018.

[HW73] Carl Hierholzer and Chr Wiener. Über die möglichkeit, einen linienzug ohne wiederholung und ohne unterbrechung zu umfahren. *Mathematische Annalen*, 6(1):30–32, 1873.

[HW03] Alan J Hoffman and Helmut W Wielandt. The variation of the spectrum of a normal matrix. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 118–120. World Scientific, 2003.

[IZ80] Claude Itzykson and J-B Zuber. The planar approximation. II. *Journal of Mathematical Physics*, 21(3):411–421, 1980.

[JEH15] Kishore Jaganathan, Yonina C Eldar, and Babak Hassibi. Phase retrieval: An overview of recent developments. *arXiv preprint arXiv:1510.07713*, 2015.

[JFHK94] Viktor K Jirsa, R Friedrich, Hermann Haken, and JA Scott Kelso. A theoretical model of phase transitions in the human brain. *Biological cybernetics*, 71(1):27–35, 1994.

[JGH18] Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: convergence and generalization in neural networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 8580–8589, 2018.

[JM13] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

[JMS04] Haixia Jia, Cris Moore, and Bart Selman. From spin glasses to hard satisfiable formulas. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 199–210. Springer, 2004.

[JOB10] Rodolphe Jenatton, Guillaume Obozinski, and Francis Bach. Structured sparse principal component analysis. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 366–373, 2010.

[Joh01] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29:295–327, 2001.

[Kab08a] Yoshiyuki Kabashima. Inference from correlated patterns: a unified theory for perceptron learning and linear vector channels. In *Journal of Physics: Conference Series*, volume 95, page 012001. IOP Publishing, 2008.

[Kab08b] Yoshiyuki Kabashima. An integral formula for large random rectangular matrices and its application to analysis of linear vector channels. In *2008 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops*, pages 620–624. IEEE, 2008.

[Kaw16] Kenji Kawaguchi. Deep learning without poor local minima. In *Advances in Neural Information Processing Systems*, pages 586–594, 2016.

[KKM+16] Yoshiyuki Kabashima, Florent Krzakala, Marc Mézard, Ayaka Sakata, and Lenka Zdeborová. Phase transitions and sample complexity in Bayes-optimal matrix factorization. *IEEE Transactions on information theory*, 62(7):4228–4265, 2016.

[Kle13] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[KMM+13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

[KMS+12] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012.

[KMTZ14] Florent Krzakala, Andre Manoel, Eric W Tramel, and Lenka Zdeborová. Variational free energies for compressed sensing. In *2014 IEEE International Symposium on Information Theory*, pages 1499–1503. IEEE, 2014.

[KRFU14] Ulugbek S Kamilov, Sundeep Rangan, Alyson K Fletcher, and Michael Unser. Approximate message passing with consistent parameter estimation and applications to sparse learning. *IEEE Transactions on Information Theory*, 60(5):2969–2985, 2014.

[KTJ76] John M Kosterlitz, David J Thouless, and Raymund C Jones. Spherical model of a spin-glass. *Physical Review Letters*, 36(20):1217, 1976.

[KU04] Yoshiyuki Kabashima and Shinsuke Uda. A BP-based algorithm for performing Bayesian inference in large perceptron-type networks. In *International Conference on Algorithmic Learning Theory*, pages 479–493. Springer, 2004.

[Kur91] J Kurchan. Replica trick to calculate means of absolute values: applications to stochastic equations. *Journal of Physics A: Mathematical and General*, 24(21):4969, 1991.

[KXZ16] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75. IEEE, 2016.

[LAL19] Wangyu Luo, Wael Alghamdi, and Yue M Lu. Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Transactions on Signal Processing*, 67(9):2347–2356, 2019.

[Lap74] Pierre Simon Laplace. Mémoire sur la probabilité de causes par les évenements. *Mémoire de l'académie royale des sciences*, 1774.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[LK17] Haihao Lu and Kenji Kawaguchi. Depth creates no bad local minima. *arXiv preprint arXiv:1702.08580*, 2017.

[LKZ15] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse PCA. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1635–1639. IEEE, 2015.

[LKZ17] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

[LL20] Yue M Lu and Gen Li. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, 9(3):507–541, 2020.

[LM19] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3-4):859–929, 2019.

[LML+17] Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017.

[LNV18] Giacomo Livan, Marcel Novaes, and Pierpaolo Vivo. *Introduction to random matrices: theory and practice*, volume 26. Springer, 2018.

[LS16] Ji Oon Lee and Kevin Schnelli. Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *The Annals of Applied Probability*, 26(6):3786–3839, 2016.

[LSL19] Carlo Lucibello, Luca Saglietti, and Yue Lu. Generalized approximate survey propagation for high-dimensional estimation. In *International Conference on Machine Learning*, pages 4173–4182, 2019.

[LXB19] Shuyang Ling, Ruitu Xu, and Afonso S Bandeira. On the landscape of synchronization networks: a perspective from nonconvex optimization. *SIAM Journal on Optimization*, 29(3):1879–1907, 2019.

[LXT+18] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 6391–6401, 2018.

[Mai07] Mylène Maida. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electronic Journal of Probability*, 12:1131–1150, 2007.

[Mat94] A Matytsin. On the large-N limit of the Itzykson-Zuber integral. *Nuclear Physics B*, 411(2-3):805–820, 1994.

[Max60] James Clerk Maxwell. Illustrations of the dynamical theory of gases. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 19-20, 1860.

[MBM18] Song Mei, Yu Bai, and Andrea Montanari. The landscape of empirical risk for nonconvex losses. *The Annals of Statistics*, 46(6A):2747–2774, 2018.

[McC18] Peter McCullagh. *Generalized linear models*. Routledge, 2018.

[McK21a] Benjamin McKenna. Complexity of bipartite spherical spin glasses. *arXiv preprint arXiv:2105.05043*, 2021.

[McK21b] Benjamin McKenna. Large deviations for extreme eigenvalues of deformed Wigner random matrices. *Electronic Journal of Probability*, 26:1–37, 2021.

[MDX+21] Junjie Ma, Rishabh Dudeja, Ji Xu, Arian Maleki, and Xiaodong Wang. Spectral method for phase retrieval: an expectation propagation perspective. *IEEE Transactions on Information Theory*, 67(2):1332–1355, 2021.

[Meh04] Madan Lal Mehta. *Random matrices*, volume 142. Elsevier, 2004.

[Méz89] Marc Mézard. The space of interactions in neural networks: Gardner's computation with the cavity method. *Journal of Physics A: Mathematical and General*, 22(12):2181, 1989.

[Méz15] Marc Mézard. Cavity method: message-passing from a physics perspective. *Statistical Physics, Optimization, Inference, and Message-Passing Algorithms: Lecture Notes of the Les Houches School of Physics: Special Issue, October 2013*, page 95, 2015.

[Méz17] Marc Mézard. Mean-field message-passing equations in the Hopfield model and its generalizations. *Physical Review E*, 95(2):022117, 2017.

[Min01] Thomas P Minka. Expectation propagation for approximate Bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[Mio17] Léo Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017.

[MKMZ17] Andre Manoel, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Multi-layer generalized linear estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2098–2102. IEEE, 2017.

[MKUZ20] Francesca Mignacco, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification. *Advances in Neural Information Processing Systems*, 33, 2020.

[MM09] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[MM11] Cristopher Moore and Stephan Mertens. *The nature of computation*. OUP Oxford, 2011.

[MM19] Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. *Foundations of Computational Mathematics*, 19(3):703–773, 2019.

[Mon95] Rémi Monasson. Structural glass transition and the entropy of the metastable states. *Physical review letters*, 75(15):2847, 1995.

[Mon21] Andrea Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. *SIAM Journal on Computing*, pages FOCS19–1, 2021.

[MP67] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967.

[MP69] Marvin Minsky and Seymour Papert. Perceptron: an introduction to computational geometry. *The MIT Press, Cambridge, expanded edition*, 19(88):2, 1969.

[MP92] German Mato and Nestor Parga. Generalization properties of multilayered neural networks. *Journal of Physics A: Mathematical and General*, 25(19):5047, 1992.

[MP17] Junjie Ma and Li Ping. Orthogonal AMP. *IEEE Access*, 5:2020–2033, 2017.

[MPR94a] Enzo Marinari, Giorgio Parisi, and Felix Ritort. Replica field theory for deterministic models: I. binary sequences with low autocorrelation. *Journal of Physics A: Mathematical and General*, 27(23):7615, 1994.

[MPR94b] Enzo Marinari, Giorgio Parisi, and Felix Ritort. Replica field theory for deterministic models. II. a non-random spin glass with glassy behaviour. *Journal of Physics A: Mathematical and General*, 27(23):7647, 1994.

[MPV86] M Mézard, G Parisi, and M. A Virasoro. SK model: The replica solution without replicas. *Europhysics Letters (EPL)*, 1(2):77–82, jan 1986.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[MS14] Satya N Majumdar and Grégory Schehr. Top eigenvalue of a random matrix: large deviations and third order phase transition. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(1):P01012, 2014.

[MSWW63] John Willard Milnor, Michael Spivak, Robert Wells, and Robert Wells. *Morse theory*. Princeton university press, 1963.

[MTV20] Marco Mondelli, Christos Thrampoulidis, and Ramji Venkataramanan. Optimal combination of linear and spectral estimators for generalized linear models. *arXiv preprint arXiv:2008.03326*, 2020.

[MUZ21] Francesca Mignacco, Pierfrancesco Urbani, and Lenka Zdeborová. Stochasticity helps to navigate rough landscapes: comparing gradient-descent-based algorithms in the phase retrieval problem. *Machine Learning: Science and Technology*, 2021.

[MV09] Satya N Majumdar and Massimo Vergassola. Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices. *Physical review letters*, 102(6):060601, 2009.

[MV21] Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. In *International Conference on Artificial Intelligence and Statistics*, pages 397–405. PMLR, 2021.

[MYP14] Junjie Ma, Xiaojun Yuan, and Li Ping. Turbo compressed sensing with partial DFT sensing matrix. *IEEE Signal Processing Letters*, 22(2):158–161, 2014.

[MZ95a] Rémi Monasson and Riccardo Zecchina. Learning and generalization theories of large committee-machines. *Modern Physics Letters B*, 9(30):1887–1897, 1995.

[MZ95b] Rémi Monasson and Riccardo Zecchina. Weight space structure and internal representations: a direct approach to learning and generalization in multilayer neural networks. *Physical review letters*, 75(12):2432, 1995.

[MZ20] Stefano Sarao Mannelli and Lenka Zdeborová. Thresholds of descending algorithms in inference problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(3):034004, 2020.

[Nik30] Otton Nikodym. Sur une généralisation des intégrales de MJ Radon. *Fundamenta Mathematicae*, 15(1):131–179, 1930.

[Nis01] Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*, volume 111. Clarendon Press, 2001.

[NJS15] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015.

[NT97] Kazuo Nakanishi and Hajime Takayama. Mean-field theory for a spin-glass model of neural networks: TAP free energy and the paramagnetic to spin-glass transition. *Journal of Physics A: Mathematical and General*, 30(23):8085, 1997.

[NW72] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

[OCW16] Manfred Opper, Burak Cakmak, and Ole Winther. A theory of solving TAP equations for Ising models with general invariant random matrices. *Journal of Physics A: Mathematical and Theoretical*, 49(11):114002, 2016.

[OS01] Manfred Opper and David Saad. *Advanced mean field methods: Theory and practice*. MIT press, 2001.

[OW96] Manfred Opper and Ole Winther. Mean field approach to Bayes learning in feed-forward neural networks. *Physical review letters*, 76(11):1964, 1996.

[OW01a] Manfred Opper and Ole Winther. Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling. *Physical Review E*, 64(5):056131, 2001.

[OW01b] Manfred Opper and Ole Winther. Tractable approximations for probabilistic models: The adaptive Thouless-Anderson-Palmer mean field approach. *Physical Review Letters*, 86(17):3695, 2001.

[OW05a] Manfred Opper and Ole Winther. Expectation consistent approximate inference. *Journal of Machine Learning Research*, 6(Dec):2177–2204, 2005.

[OW05b] Manfred Opper and Ole Winther. Expectation consistent free energies for approximate inference. In *Advances in Neural Information Processing Systems*, pages 1001–1008, 2005.

[Par79] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[Par80a] Giorgio Parisi. The order parameter for spin glasses: a function on the interval 0-1. *Journal of Physics A: Mathematical and General*, 13(3):1101, 1980.

[Par80b] Giorgio Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

[Pea82] Judea Pearl. *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science, 1982.

[PKCS17] Dohyung Park, Anastasios Kyrillidis, Constantine Carmanis, and Sujay Sanghavi. Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach. In *Artificial Intelligence and Statistics*, pages 65–74, 2017.

[Ple82] Timm Plefka. Convergence condition of the TAP equation for the infinite-ranged ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6):1971, 1982.

[PP95] Giorgio Parisi and Marc Potters. Mean-field equations for spin models with orthogonal interaction matrices. *Journal of Physics A: Mathematical and General*, 28(18):5267, 1995.

[PS21] Vanessa Piccolo and Dominik Schröder. Analysis of one-hidden-layer neural networks via the resolvent method. *arXiv preprint arXiv:2105.05115*, 2021.

[PSC14a] Jason T Parker, Philip Schniter, and Volkan Cevher. Bilinear generalized approximate message passing—part i: Derivation. *IEEE Transactions on Signal Processing*, 62(22):5839–5853, 2014.

[PSC14b] Jason T Parker, Philip Schniter, and Volkan Cevher. Bilinear generalized approximate message passing—Part II: Applications. *IEEE Transactions on Signal Processing*, 62(22):5854–5867, 2014.

[PW19] Jeffrey Pennington and Pratik Worah. Nonlinear random matrix theory for deep learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124005, 2019.

[PWBM16] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of PCA for spiked random matrices and synchronization. *arXiv preprint arXiv:1609.05573*, 2016.

[Ran10] Sundeep Rangan. Estimation with random linear mixing, belief propagation and compressed sensing. In *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2010.

[Ran11] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.

[RBABC19] Valentina Ros, Gérard Ben Arous, Giulio Biroli, and Chiara Cammarota. Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions. *Physical Review X*, 9(1):011003, 2019.

[Ree17] Galen Reeves. Additivity of information in multilayer networks via additive gaussian noise transforms. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1064–1070. IEEE, 2017.

[RF12] Sundeep Rangan and Alyson K Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1246–1250. IEEE, 2012.

[RHW86] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[Ric44] Stephen O Rice. Mathematical analysis of random noise. *The Bell System Technical Journal*, 23(3):282–332, 1944.

[Ros57] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton*. Cornell Aeronautical Laboratory, 1957.

[RPD18] Galen Reeves, Henry D Pfister, and Alex Dytso. Mutual information as a function of matrix SNR for linear gaussian channels. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1754–1758. IEEE, 2018.

[RPP08] David Reich, Alkes L Price, and Nick Patterson. Principal component analysis of genetic data. *Nature genetics*, 40(5):491–492, 2008.

[RSF17] Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Vector approximate message passing. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1588–1592. IEEE, 2017.

[RTWZ01] Federico Ricci-Tersenghi, Martin Weigt, and Riccardo Zecchina. Simplest random k-satisfiability problem. *Physical Review E*, 63(2):026702, 2001.

[RXZ19] Galen Reeves, Jiaming Xu, and Ilias Zadik. All-or-nothing phenomena: From single-letter to high dimensions. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 654–658. IEEE, 2019.

[San58] Ivan N Sanov. On the probability of large deviations of random variables. Technical report, North Carolina State University. Dept. of Statistics, 1958.

[SB95] Jack W Silverstein and ZD Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, 54(2):175–192, 1995.

[SC95] Jack W Silverstein and Sang-Il Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):295–309, 1995.

[SC16] Daniel Soudry and Yair Carmon. No bad local minima: Data independent training error guarantees for multilayer neural networks. *arXiv preprint arXiv:1605.08361*, 2016.

[Sch42] IJ Schoenberg. Positive definite functions on spheres. *Duke Mathematical Journal*, 9(1):96–108, 1942.

[Sch93] Henry Schwarze. Learning a rule in a multilayer neural network. *Journal of Physics A: Mathematical and General*, 26(21):5781, 1993.

[Sch16] Christophe Schülke. *Statistical physics of linear and bilinear inference problems*. PhD thesis, Sorbonne Paris Cité, 2016.

[SEC+15] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[Sel21] Mark Sellke. Optimizing mean field spin glasses with external field. *arXiv preprint arXiv:2105.03506*, 2021.

[SH92] Henry Schwarze and John Hertz. Generalization in a large committee machine. *EPL (Europhysics Letters)*, 20(4):375, 1992.

[SH93] Henry Schwarze and John Hertz. Generalization in fully connected committee machines. *EPL (Europhysics Letters)*, 21(7):785, 1993.

[Sha48] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[SIH10] Nen Saito, Yukito Iba, and Koji Hukushima. Multicanonical sampling of rare events in random matrices. *Physical Review E*, 82(3):031142, 2010.

[Sil95] Jack W Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995.

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.

[SK08] Takashi Shinzato and Yoshiyuki Kabashima. Perceptron capacity revisited: classification ability for correlated patterns. *Journal of Physics A: Mathematical and Theoretical*, 41(32):324013, 2008.

[SKZ14] Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the Bethe Hessian. In *Advances in Neural Information Processing Systems*, pages 406–414, 2014.

[SLJ+15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[SMBC+19] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, and Lenka Zdeborová. Who is afraid of big bad minima? analysis of gradient-flow in spiked matrix-tensor models. In *Advances in Neural Information Processing Systems*, pages 8679–8689, 2019.

[SMBC+20a] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Complex dynamics in simple neural networks: Understanding gradient flow in phase retrieval. In *Advances in Neural Information Processing Systems*, 2020.

[SMBC+20b] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Marvels and pitfalls of the Langevin algorithm in noisy high-dimensional inference. *Physical Review X*, 10(1):011057, 2020.

[SMKUZ19] Stefano Sarao Mannelli, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Passed & spurious: Descent algorithms and local minima in spiked matrix-tensor models. In *International Conference on Machine Learning*, pages 4333–4342, 2019.

[SQW18] Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.

[SR14] Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2014.

[SRF16] Philip Schniter, Sundeep Rangan, and Alyson K Fletcher. Vector approximate message passing for the generalized linear model. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1525–1529. IEEE, 2016.

[SS95] David Saad and Sara A Solla. Online learning in soft committee machines. *Physical Review E*, 52(4):4225, 1995.

[SS18] Itay Safran and Ohad Shamir. Spurious local minima are common in two-layer relu neural networks. In *International Conference on Machine Learning*, pages 4433–4441. PMLR, 2018.

[SST92] Sebastian Seung, Haim Sompolinsky, and Naftali Tishby. Statistical mechanics of learning from examples. *Physical Review A*, 45(8):6056, 1992.

[Ste89] Daniel L Stein. Spin glasses. *Scientific American*, 261(1):52–61, 1989.

[STS90] Haim Sompolinsky, Naftali Tishby, and H Sebastian Seung. Learning from examples in large neural networks. *Physical Review Letters*, 65(13):1683, 1990.

[Sub17a] Eliran Subag. The complexity of spherical p-spin models—a second moment approach. *The Annals of Probability*, 45(5):3385–3450, 2017.

[Sub17b] Eliran Subag. The geometry of the Gibbs measure of pure spherical spin glasses. *Inventiones mathematicae*, 210(1):135–209, 2017.

[Sub21] Eliran Subag. Following the ground states of Full-RSB spherical spin glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021.

[SZ17] Eliran Subag and Ofer Zeitouni. The extremal process of critical points of the pure p-spin spherical spin glass model. *Probability theory and related fields*, 168(3):773–820, 2017.

[Tal03] Michel Talagrand. *Spin glasses: a challenge for mathematicians: cavity and mean field models*, volume 46. Springer Science & Business Media, 2003.

[Tal06] Michel Talagrand. The Parisi formula. *Annals of mathematics*, pages 221–263, 2006.

[TAP77] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of 'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

[TCVS13] Antonia M Tulino, Giuseppe Caire, Sergio Verdu, and Shlomo Shamai. Support recovery with sparsely sampled free random matrices. *IEEE Transactions on Information Theory*, 59(7):4243–4271, 2013.

[TK20] Takashi Takahashi and Yoshiyuki Kabashima. Macroscopic analysis of vector approximate message passing in a model mismatch setting. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1403–1408. IEEE, 2020.

[TMC+16] Eric W Tramel, Andre Manoel, Francesco Caltagirone, Marylou Gabrié, and Florent Krzakala. Inferring sparsity: Compressed sensing using generalized restricted Boltzmann machines. In *2016 IEEE Information Theory Workshop (ITW)*, pages 265–269. IEEE, 2016.

[TUK06] Koujin Takeda, Shinsuke Uda, and Yoshiyuki Kabashima. Analysis of CDMA systems that are characterized by eigenvalue spectrum. *EPL (Europhysics Letters)*, 76(6):1193, 2006.

[TV04] Antonia M Tulino and Sergio Verdú. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1):1–182, 2004.

[TW94] Craig A Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

[UE88] Michael Unser and Murray Eden. Maximum likelihood estimation of liner signal parameters for poisson processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(6):942–945, 1988.

[Vap98] Vladimir Vapnik. *Statistical learning theory. 1998*. Wiley, New York, 1998.

[Vap13] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[VDG21] Lorenzo Valzania, Jonathan Dong, and Sylvain Gigan. Accelerating ptychographic reconstructions using spectral initializations. *Optics Letters*, 46(6):1357–1360, 2021.

[VDN92] Dan V Voiculescu, Ken J Dykema, and Alexandru Nica. *Free random variables*. American Mathematical Soc., 1992.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[Vil18] Soledad Villar. Generative models are the new sparsity? https://solevillar.github.io/2018/03/28/SUNLayer.html, 2018.

[VMB07] Pierpaolo Vivo, Satya N Majumdar, and Oriol Bohigas. Large deviations of the maximum eigenvalue in Wishart random matrices. *Journal of Physics A: Mathematical and Theoretical*, 40(16):4317, 2007.

[Voi87] Dan Voiculescu. Multiplication of certain non-commuting random variables. *Journal of Operator Theory*, pages 223–235, 1987.

[VW00] JJM Verbaarschot and T Wettig. Random matrix theory and chiral symmetry in qcd. *Annual Review of Nuclear and Particle Science*, 50(1):343–410, 2000.

[Wal18] Irene Waldspurger. Phase retrieval with random gaussian sensing vectors by alternating projections. *IEEE Transactions on Information Theory*, 64(5):3301–3312, 2018.

[Wei78] Don Weingarten. Asymptotic behavior of group integrals in the limit of infinite rank. *Journal of Mathematical Physics*, 19(5):999–1001, 1978.

[Wey12] Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

[Wey49] Hermann Weyl. Inequalities between the two kinds of eigenvalues of a linear transformation. *Proceedings of the National Academy of Sciences of the United States of America*, 35(7):408, 1949.

[Wig55] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955.

[Wis28] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52, 1928.

[WRB93] Timothy LH Watkin, Albrecht Rau, and Michael Biehl. The statistical mechanics of learning a rule. *Reviews of Modern Physics*, 65(2):499, 1993.

[XRV17] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[ZBH+16] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *Communications of the ACM*, 64, 11 2016.

[ZBK16] Junan Zhu, Dror Baron, and Florent Krzakala. Performance limits for noisy multimeasurement vector problems. *IEEE Transactions on Signal Processing*, 65(9):2444–2454, 2016.

[ZCM+16] Fucai Zhang, Bo Chen, Graeme R Morrison, Joan Vila-Comamala, Manuel Guizar-Sicairos, and Ian K Robinson. Phase retrieval by coherent modulation imaging. *Nature communications*, 7(1):1–8, 2016.

[Zde20] Lenka Zdeborová. Understanding deep learning is also a job for physicists. *Nature Physics*, 16:602–604, 2020.

[ZDZS10] Lei Zhang, Weisheng Dong, David Zhang, and Guangming Shi. Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern recognition*, 43(4):1531–1549, 2010.

[ZJZ03] Paul Zinn-Justin and Jean-Bernard Zuber. On some integrals over the $\mathcal{U}(N)$ unitary group and their large$-N$ limit. *Journal of Physics A: Mathematical and General*, 36(12):3173, 2003.

[ZK16] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

[Zub18] Jean-Bernard Zuber. Horn's problem and Harish-Chandra's integrals. Probability density functions. *Annales de l'Institut Henri Poincaré D*, 5(3):309–338, 2018.

[ZZY20] Qiuyun Zou, Haochuan Zhang, and Hongwen Yang. Multi-Layer Bilinear Generalized Approximate Message Passing. *arXiv preprint arXiv:2007.00436*, 2020.

Appendix A

Technicalities of the Plefka-Georges-Yedidia expansion

A.1 Order 4 of the expansion for a spherical model

We start from eq. (2.7d), that we consider at $\eta = 0$:

$$n\partial_\eta^4 \Phi_{\mathbf{J}} = \langle U^4\rangle_0 - 3\langle U^2\rangle_0^2 - 3\sum_{i=1}^n \partial_\eta^2\lambda_i\langle U^2(x_i - m_i)\rangle_0 - \frac{3}{2}\sum_{i=1}^n \partial_\eta^2\gamma_i\langle U^2(x_i^2 - m_i^2 - v_i)\rangle_0. \quad \text{(A.1)}$$

For simplicity we will denote $\tilde{x}_i \equiv (x_i - m_i)$, so that at $\eta = 0$ the $\{\tilde{x}_i\}$ variables are Gaussian variables with mean $\langle\tilde{x}_i\rangle = 0$ and covariance $\langle\tilde{x}_i\tilde{x}_j\rangle = \delta_{ij}v_i$. In particular eq. (2.10) becomes:

$$U(\eta = 0, \mathbf{J}) = -\frac{1}{2}\sum_{i\neq j} J_{ij}\tilde{x}_i\tilde{x}_j.$$

From our calculation at order 2 we obtain the following relation that we can represent diagrammatically:

$$-3\langle U^2\rangle_0^2 = -\frac{3}{4}\Big[\sum_{i\neq j} J_{ij}^2 v_i v_j\Big]^2 = -\frac{3n}{4}\Big[$$  $$\Big]^2. \quad \text{(A.2)}$$

We now turn to the next term:

$$\begin{aligned}
&-\frac{3}{2}\sum_{i=1}^n \partial_\eta^2\gamma_i\langle U^2(x_i^2 - m_i^2 - v_i)\rangle_0 - 3\sum_{i=1}^n \partial_\eta^2\lambda_i\langle U^2(x_i - m_i)\rangle_0 \\
&= -\frac{3}{2}\sum_{i=1}^n \partial_\eta^2\gamma_i\langle U^2(\tilde{x}_i^2 - v_i)\rangle_0 - 3\sum_{i=1}^n \langle U^2\tilde{x}_i(\partial_\eta^2\lambda_i + m_i\partial_\eta^2\gamma_i)\rangle_0, \\
&\overset{(a)}{=} -\frac{3}{2}\sum_{i=1}^n \partial_\eta^2\gamma_i\langle U^2(\tilde{x}_i^2 - v_i)\rangle_0 - 3n\sum_{i=1}^n \langle U^2\tilde{x}_i\frac{\partial(\partial_\eta^2\Phi_{\mathbf{J}})}{\partial m_i}\rangle_0 \overset{(b)}{=} -\frac{3}{2}\sum_{i=1}^n \partial_\eta^2\gamma_i\langle U^2(\tilde{x}_i^2 - v_i)\rangle_0 + \mathcal{O}_n(1).
\end{aligned}$$

In (a) we used the Maxwell equations (2.9), while in (b) we made use of the fact that the order 2 of the free entropy does not depend on the m_i variables. We obtain for the remaining term:

$$-\frac{3}{2}\sum_{i=1}^n \partial_\eta^2\gamma_i\langle U^2(\tilde{x}_i^2 - v_i)\rangle_0 = \frac{3}{2}\Big[\sum_{i\neq j} J_{ij}^2 v_i v_j\Big]^2 - 3\sum_{i\neq j} J_{ij}^2 v_j\langle U^2\tilde{x}_i^2\rangle_0,$$

in which we used the Maxwell relations (2.9) to compute $\partial_\eta^2\gamma_i$. To compute $\langle U^2\tilde{x}_i^2\rangle_0$, we expand:

$$\langle U^2\tilde{x}_i^2\rangle_0 = \frac{1}{4}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2} J_{i_1 j_1}J_{i_2 j_2}\langle\tilde{x}_i^2\tilde{x}_{i_1}\tilde{x}_{j_1}\tilde{x}_{i_2}\tilde{x}_{j_2}\rangle_0.$$

(A) (B)

FIGURE A.1: Different types of diagrams of indices appearing in $\langle U^4 \rangle_0$.

We can then use Wick's theorem to simplify the average. There are two types of contractions (or pairings) that appear:

- Contractions that do not mix the indices i_1, j_1, i_2, j_2 with i. There are 2 such possible pairings and they give rise to the diagram $\Big[n$  $\Big]^2$.
- Contractions that mix these indices with i. There are all equivalent and there are 8 of them, each giving the diagram n  .

In the end, we reach:

$$\langle U^2 \tilde{x}_i^2 \rangle_0 = \frac{v_i}{2} \sum_{k \neq l} J_{kl}^2 v_k v_l + 2 \sum_{k(\neq i)} J_{ik}^2 v_i^2 v_k.$$

We can finally compute the term we were seeking:

$$-\frac{3}{2} \sum_{i=1}^{n} \partial_\eta^2 \gamma_i \langle U^2 (\tilde{x}_i^2 - v_i) \rangle_0 = -6n \sum_{\substack{i,j,k \\ \text{pairwise distincts}}} J_{ij}^2 J_{ik}^2 v_i^2 v_j v_k = -6$$  . (A.3)

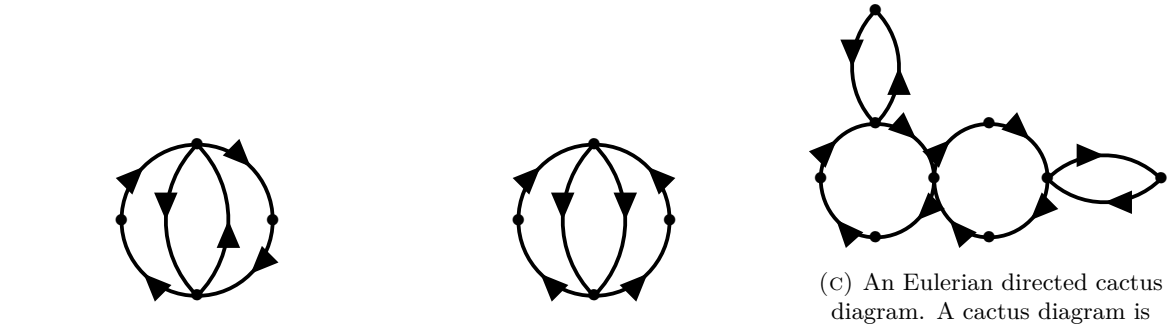
Note that in this last equation we could add the hypothesis that $j \neq k$. Indeed the term $j = k$ would give rise to the diagram n  , which is negligible, since for every $i \neq j$ one has $J_{ij} = \mathcal{O}(n^{-1/2})$ as a consequence of rotational invariance (cf. Model S). We finally turn to the computation of $\langle U^4 \rangle_0$:

$$\langle U^4 \rangle_0 = \frac{1}{16} \prod_{\alpha=0}^{3} \Big[\sum_{i_\alpha \neq j_\alpha} J_{i_\alpha j_\alpha} \Big] \Big\langle \prod_{\alpha=0}^{3} \tilde{x}_{i_\alpha} \tilde{x}_{j_\alpha} \Big\rangle_0.$$

The possible contractions arising from Wick's theorem yield several contributions, that we can also represent by diagrams. Note that these diagrams are very different from the diagrams that we described for instance in Fig. 2.1, and are merely a way to visualize the contractions in Wick's theorem. The first column contains the i_α indices and the second contains the j_α. Note that we always have $i_\alpha \neq j_\alpha$. The two different types of contractions are represented as Fig. A.1a and Fig. A.1b. They are 12 possible contractions of the type of Fig. A.1a and 48 of Fig. A.1b. We also take into account that in the pairings of Fig. A.1b indices are not all necessarily pairwise distinct. Discarding terms that are $\mathcal{O}(n)$, we finally reach:

$$\langle U^4 \rangle_0 = \frac{3}{4} \Big[\sum_{i \neq j} J_{ij}^2 v_i v_j \Big]^2 + 6 \sum_{\substack{i,j,k \\ \text{pairwise distincts}}} J_{ij}^2 J_{ik}^2 v_i^2 v_j v_k + 3 \sum_{\substack{i_0, i_1, i_2, i_3 \\ \text{pairwise distincts}}} J_{i_0 i_1} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_0} v_{i_0} v_{i_1} v_{i_2} v_{i_3},$$

$$= \frac{3}{4} \Big[n$$  $\Big]^2 + 6n$  $. + 3n$  . (A.4)



(A) A directed Eulerian diagram for Hermitian matrices.

(B) A different Eulerian diagram, but with the same undirected edges.

(C) An Eulerian directed cactus diagram. A cactus diagram is Eulerian if and only if all its simple cycles are Eulerian.

FIGURE A.2: Diagrams similar to the ones of Fig. 2.2, but for Hermitian matrices. Note that the diagrams of Fig. A.2a and Fig. A.2b are different because of the different directions of the edges, but that both are Eulerian.

Finally, combining eqs. (A.2),(A.3),(A.4) to plug them into eq. (A.1), we reach:

$$\frac{1}{4!}\frac{\partial^4 \Phi_J}{\partial \eta^4} = \frac{1}{8n} \sum_{\substack{i_0,i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_0 i_1} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_0} v_{i_0} v_{i_1} v_{i_2} v_{i_3} + \mathcal{O}_n(1) = \frac{1}{8}$$  $$+ \mathcal{O}_n(1),$$

which is what we wanted to show !

A.2 Generalizations of the diagrammatics

We detail here some extensions of the results of Section 2.4. First, in Sec. A.2.1, we explain how to transpose these results to Hermitian matrix models, and in Sec. A.2.2 we show how to extend some of them to diagrams of diverging size (as $n \to \infty$).

A.2.1 Hermitian matrix model

We can generalize the results of Sec. 2.4 to the Hermitian matrix model described by Model S, i.e. when $\beta = 2$. Note that the diagrams for Hermitian matrices are *directed*, as $J_{ij} = \overline{J_{ji}}$. We describe examples of such diagrams in Fig. A.2. E.g. the diagrams of Fig. A.2a,A.2b are respectively equal to:

$$\frac{1}{n} \sum_{\substack{i_1,\cdots,i_4 \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1} |J_{i_2 i_4}|^2 \qquad \text{and} \qquad \frac{1}{n} \sum_{\substack{i_1,\cdots,i_4 \\ \text{pairwise distincts}}} J_{i_1 i_2} \overline{J_{i_2 i_3} J_{i_3 i_4}} J_{i_4 i_1} J_{i_2 i_4}^2.$$

In the complex case, an *Eulerian* graph is defined as a graph in which one can construct a cyclic path (following the directions of the edges) that visits each edge exactly once. Note that a *simple cycle* is defined such that the arrows on its edges themselves form a cycle, like the constituent cycles of Fig. A.2c. We describe the main results we get, using the same kind of techniques as used in Sec. 2.4:

(*i*) Only Eulerian diagrams contribute in the $n \to \infty$ limit.

(*ii*) Consider a simple cycle $\mathcal{C}_p$ with p vertices. Then this diagram converges as $n \to \infty$, a.s. and in L^2 norm, to the free cumulant $c_p(\rho)$, as in the real case. More precisely:

$$\lim_{n\to\infty} \mathbb{E}_{\mathbf{U}} \left| \frac{1}{n} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} (\mathbf{U}\mathbf{D}\mathbf{U}^\dagger)_{i_1 i_2} (\mathbf{U}\mathbf{D}\mathbf{U}^\dagger)_{i_2 i_3} \cdots (\mathbf{U}\mathbf{D}\mathbf{U}^\dagger)_{i_p i_1} - c_p(\rho) \right|^2 = 0.$$

(*iii*) Any Eulerian strongly irreducible diagram that is not a simple cycle will be negligible in the $n \to \infty$ limit (in L^2 norm).

(*iv*) Any Eulerian cactus diagram (like in Fig. A.2c) will converge in L^2 to the products of the free cumulants of ρ corresponding to each one of its constituent simple cycles.

These results are straightforward generalizations of the ones obtained for real matrices in Sec. 2.4. For completeness, we describe how to show a weaker version of (*ii*), and leave the other statements to the reader. As before, by unitary invariance we can assume that $(i_1, \cdots, i_p) = (1, , \cdots, p)$, and we can apply the results of [GM05] to obtain a similar equation to eq. (2.49):

$$\begin{aligned} L_p &\equiv \lim_{n\to\infty} n^{p-1} \mathbb{E}_{\mathbf{U}} \left[(\mathbf{U}\mathbf{D}\mathbf{U}^\dagger)_{12} \cdots (\mathbf{U}\mathbf{D}\mathbf{U}^\dagger)_{p1} \right], \\ &= \lim_{n\to\infty} \frac{1}{n} \prod_{l=1}^{p} \left(\frac{\partial}{\partial b_i} + i \frac{\partial}{\partial c_i} \right) \left[\exp\left\{ n \sum_{n=1}^{\infty} \frac{c_n(\rho)}{n} \mathrm{Tr}\left[\mathbf{M}(\mathbf{b},\mathbf{c})^n \right] \right\} \right]_{\mathbf{b},\mathbf{c}=0}, \end{aligned} \tag{A.5}$$

with now the matrix $\mathbf{M}(\mathbf{b},\mathbf{c})$ defined as:

$$\mathbf{M}(\mathbf{b},\mathbf{c}) \equiv \frac{1}{2} \begin{pmatrix} 0 & b_1 + ic_1 & 0 & \cdots & 0 & b_p - ic_p \\ b_1 - ic_1 & 0 & b_2 + ic_2 & \cdots & 0 & 0 \\ 0 & b_2 - ic_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & b_{p-1} + ic_{p-1} \\ b_p + ic_p & 0 & 0 & \cdots & b_{p-1} - ic_{p-1} & 0 \end{pmatrix}.$$

We have $[\partial_{b_i} + i\partial_{c_i}]\mathbf{M}(\mathbf{b},\mathbf{c}) = \mathbf{F}_{i+1,i}$, in which $(\mathbf{F}_{a,b})_{ll'} \equiv \delta_{al}\delta_{bl'}$ are elementary non-symmetric matrices. In the exact same way as in Sec. 2.4.1, the dominant contribution in eq. (A.5) will be given by differentiating a single time the exponential term, and creating a cycle with the matrices $\mathbf{F}_{i+1,i}$. Note that contrary to the symmetric case of Sec. 2.4.1, here *only the directed cycle will contribute*, whereas both possible directions of the cycle contributed in eq. (2.49). Indeed, the cycles in terms of the matrices $\{\mathbf{F}_{a,b}\}$ have to be directed in order to yield a non-zero contribution:

$$\mathrm{Tr}\left[\mathbf{F}_{1,2}\mathbf{F}_{2,1}\mathbf{F}_{1,3}\mathbf{F}_{3,2}\mathbf{F}_{2,1}\right] \neq 0, \quad \text{while} \quad \mathrm{Tr}\left[\mathbf{F}_{1,2}\mathbf{F}_{2,1}\mathbf{F}_{2,3}\mathbf{F}_{3,2}\mathbf{F}_{2,1}\right] = 0.$$

Thus we reach:

$$L_p = \sum_{k=p}^{\infty} c_k(\rho) \mathrm{Tr}\left[\left(\mathbf{F}_{1,p}\mathbf{F}_{p,p-1}\cdots\mathbf{F}_{2,1} \right) \mathbf{M}(0,0)^{n-p} \right] = c_p(\rho).$$

In order to get L^2 concentration of the simple cycle on the free cumulant, one can then exactly repeat the arguments of Sec. 2.4.3.

A.2.2 A note on the expectation of diagrams of diverging size

Although it is not directly useful in our PGY expansions, another side question one can ask on the behavior of these diagrams is: how do diagrams that have a number of edges that diverges with n behave in the $n \to \infty$ limit ? In all of Sec. 2.4 we only considered diagrams of finite size. The behavior of HCIZ-type integrals with a matrix with diverging rank (as opposed to the finite-rank case) has first been analyzed rigorously in [GM05] for a rank $\mathcal{O}(n^{1/2-\epsilon})$, and then generalized in [CŚ07] for ranks $\mathcal{O}(n)$. We recall the main result of [CŚ07]:

Theorem A.1 (*Collins-Śniadyc*)

Let $\mathbf{A}_n, \mathbf{B}_n$ be diagonal real matrices of size n. Assume that the rank $m(n)$ of $\mathbf{A}_n$ is such that $m = m(n) = \mathcal{O}(n)$, and denote $a_{1,n} \geq \cdots \geq a_{m,n}$ the eigenvalues of $\mathbf{A}_n$. Assume that the spectral measure of $\mathbf{B}_n$ converges a.s. and in the weak sense to a probability measure ρ_B, and that all elements of $\mathbf{A}_n$ are bounded by a constant independent of n. Then one has:

$$\frac{1}{nm} \ln \int_{\mathcal{U}(n)} \mathcal{D}\mathbf{U}\, e^{n\mathrm{Tr}[\mathbf{A}_n \mathbf{U} \mathbf{B}_n \mathbf{U}^\dagger]} = \frac{2}{m} \mathrm{Tr}[G_{\rho_B}(\mathbf{A}_n)] + \mathcal{O}_n(1).$$

A similar result holds for real orthogonal matrices:

$$\frac{1}{nm} \ln \int_{\mathcal{O}(n)} \mathcal{D}\mathbf{O}\, e^{\frac{n}{2}\mathrm{Tr}[\mathbf{A}_n \mathbf{O} \mathbf{B}_n \mathbf{O}^\intercal]} = \frac{1}{m} \mathrm{Tr}[G_{\rho_B}(\mathbf{A}_n)] + \mathcal{O}_n(1).$$

Recall that we defined the G_ρ function in Sec. 1.5 in a variational form, and showed that it can be related to the integral of the R-transform. Using Theorem A.1, the techniques of Sec. 2.4 generalize to this case.

To illustrate this last claim, we consider real symmetric matrices under Model S. We say that a sequence $\{p(n)\}$ satisfies the *bounded free cumulant property* if there exists $C > 0$ such that for all n, $|c_{p(n)}(\rho)| < C$. We state two of the results of Sec. 2.4 that can be easily generalized to the diverging size case without changing any of the arguments:

(*a*) Consider a sequence $p = p(n) = \mathcal{O}(n)$ that satisfies the bounded free cumulant property. Then one obtains the generalization of eq. (2.47):

$$n^{p(n)-1} \int_{\mathcal{O}(n)} \mathcal{D}\mathbf{O}[(\mathbf{O}\mathbf{D}\mathbf{O}^\intercal)_{12}(\mathbf{O}\mathbf{D}\mathbf{O}^\intercal)_{23} \cdots (\mathbf{O}\mathbf{D}\mathbf{O}^\intercal)_{p(n)1}] = c_{p(n)}(\rho_{\mathbf{D}}) + \mathcal{O}_n(1).$$

(*c*) Consider a cactus diagram G composed of $p(n)$ simple cycles of size $(r_1(n), \cdots, r_{p(n)}(n))$, joining at vertices. Assume that $\sum_{i=1}^{p(n)} r_i(n) = \mathcal{O}_n(n)$ and that all the sequences $r_i(n)$ satisfy the bounded free cumulant property. Then one has:

$$\mathbb{E}\, G = \Big[\prod_{i=1}^{p(n)} c_{r_i(n)}(\rho) \Big] (1 + \mathcal{O}_n(1)).$$

Other results obtained in Sec. 2.4 for finite-size diagrams might also be applicable to the diverging size case, but they are not investigated in this thesis.

A.3 PGY for extensive-rank matrix factorization

We describe here in more details the formalism we used to derive Result 3.1. Many parts of the derivation are very similar to what we described in Chapter 2, and we will refer to it when

necessary. We start from eqs. (3.8) and (3.9):

$$\begin{aligned}
n(m+p)\Phi_{\mathbf{Y},n}(\eta) &= \sum_{\mu,i}\Big[\lambda^F_{\mu i}m^F_{\mu i} + \frac{\gamma^F_{\mu i}}{2}(v^F_{\mu i}+(m^F_{\mu i})^2)\Big] + \sum_{i,l}\Big[\lambda^X_{il}m^X_{il} + \frac{\gamma^X_{il}}{2}(v^X_{il}+(m^X_{il})^2)\Big],\\
&\quad + \sum_{\mu,l}\Big[-\omega_{\mu l}g_{\mu l} - \frac{b_{\mu l}}{2}(-r_{\mu l}+g^2_{\mu l})\Big] + \ln\int P_H(\mathrm{d}\mathbf{H})\,P_F(\mathrm{d}\mathbf{F})\,P_X(\mathrm{d}\mathbf{X})\,e^{-S_{\mathrm{eff},\eta}[\mathbf{F},\mathbf{X},\mathbf{H}]},\\
S_{\mathrm{eff},\eta}[\mathbf{F},\mathbf{X},\mathbf{H}] &\equiv \sum_{\mu,i}\Big[\lambda^F_{\mu i}F_{\mu i} + \frac{\gamma^F_{\mu i}}{2}F^2_{\mu i}\Big] + \sum_{i,l}\Big[\lambda^X_{il}X_{il} + \frac{\gamma^X_{il}}{2}X^2_{il}\Big]\\
&\quad + \sum_{\mu,l}\Big[\omega_{\mu l}(iH)_{\mu l} - \frac{b_{\mu l}}{2}(iH)^2_{\mu l}\Big] + \frac{\eta}{\sqrt{n}}\sum_{\mu,i,l}(iH)_{\mu l}\,F_{\mu i}\,X_{il},\\
H_{\mathrm{eff}}[\mathbf{F},\mathbf{X},\mathbf{H}] &\equiv \frac{1}{\sqrt{n}}\sum_{\mu,i,l}(iH)_{\mu l}F_{\mu i}X_{il}.
\end{aligned}$$

A.3.1 Orders 1 and 2 in η

At order 1, we have directly:

$$\Big(\frac{\partial\Phi_{\mathbf{Y},n}}{\partial\eta}\Big)_{\eta=0} = -\frac{1}{n(m+p)}\langle H_{\mathrm{eff}}\rangle_0 = \frac{1}{n^{3/2}(m+p)}\sum_{\mu,i,l} g_{\mu l}m^F_{\mu i}m^X_{il}. \tag{A.6}$$

As in Chapter 2, we can then use "Maxwell" relations to compute the derivatives of the Lagrange parameters at $\eta=0$. For instance, for $\lambda^F_{\mu i}$ and $\gamma^F_{\mu i}$ they read:

$$\lambda^F_{\mu i} + m^F_{\mu i}\gamma^F_{\mu i} = n(m+p)\,\frac{\partial\Phi_{\mathbf{Y},n}}{\partial m^F_{\mu i}}, \qquad \gamma^F_{\mu i} = 2n(m+p)\,\frac{\partial\Phi_{\mathbf{Y},n}}{\partial v^F_{\mu i}}. \tag{A.7}$$

From it we reach easily $\partial_\eta\gamma^F_{\mu i}(\eta=0)=0$. Applying this technique to all Lagrange parameters, we can compute the operator $U_{\mathbf{Y},0}$ (defined in eq. (2.6)). For clarity of the notation, we will denote by lowercase letters *centered variables*, for instance $x_{\mu i}\equiv X_{\mu i}-m^X_{\mu i}$. We obtain:

$$U_{\mathbf{Y},0} = \frac{1}{\sqrt{n}}\sum_{\mu,i,l}[(ih)_{\mu l}f_{\mu i}x_{il} - g_{\mu l}f_{\mu i}x_{il} + (ih)_{\mu l}m^F_{\mu i}x_{il} + (ih)_{\mu l}f_{\mu i}m^X_{il}]. \tag{A.8}$$

One can then compute:

$$\begin{aligned}
\frac{1}{2}\Big(\frac{\partial^2\Phi_{\mathbf{Y},n}}{\partial\eta^2}\Big)_{\eta=0} &= \frac{1}{2n(m+p)}\langle U^2\rangle_0,\\
&= \frac{1}{2n^2(m+p)}\sum_{\mu,i,l}[-r_{\mu l}v^F_{\mu i}v^X_{il} + g^2_{\mu l}v^F_{\mu i}v^X_{il} - r_{\mu l}(m^F_{\mu i})^2v^X_{il} - r_{\mu l}v^F_{\mu i}(m^X_{il})^2].
\end{aligned} \tag{A.9}$$

A.3.2 Order 3 in η

To describe the results at order 3, we need to introduce the third cumulants of the variables $\{iH_{\mu l}, F_{\mu i}, X_{il}\}$ at $\eta=0$ (all these variables are thus independent). These cumulants are denoted $\{\kappa^{(3,H)}_{\mu l}, \kappa^{(3,F)}_{\mu i}, \kappa^{(3,X)}_{il}\}$. Using the order 3 formula of eq. (2.7), one has:

$$\frac{1}{3!}\Big(\frac{\partial^3\Phi_{\mathbf{Y},n}}{\partial\eta^3}\Big)_{\eta=0} = -\frac{1}{6n(m+p)}\langle U^3\rangle_0$$

Recall the form of the operator U at $\eta = 0$, i.e. eq. (A.8):

$$U_{\mathbf{Y},0} = \frac{1}{\sqrt{n}} \sum_{\mu,i,l} \Big[\underbrace{(ih)_{\mu l} f_{\mu i} x_{il}}_{A} + \underbrace{(-g_{\mu l} f_{\mu i} x_{il})}_{B_H} + \underbrace{(ih)_{\mu l} m^F_{\mu i} x_{il}}_{B_F} + \underbrace{(ih)_{\mu l} f_{\mu i} m^X_{il}}_{B_X} \Big]. \tag{A.10}$$

To compute $\langle U^3 \rangle_0$, we decompose $U = A + B_H + B_F + B_X$ as in the equation above. Recall that all the variables in the equation above are *centered*, so that we get easily:

$$\begin{aligned} &\langle A^3 + B_H^3 + B_F^3 + B_X^3 \rangle_0 \\ &= \frac{1}{n^{3/2}} \sum_{\mu,i,l} [\kappa^{(3,H)}_{\mu l} \kappa^{(3,F)}_{\mu i} \kappa^{(3,X)}_{il} - g^3_{\mu l} \kappa^{(3,F)}_{\mu i} \kappa^{(3,X)}_{il} + \kappa^{(3,H)}_{\mu l} (m^F_{\mu i})^3 \kappa^{(3,X)}_{il} + \kappa^{(3,H)}_{\mu l} \kappa^{(3,F)}_{\mu i} (m^X_{il})^3]. \end{aligned} \tag{A.11}$$

Using again the centering of the variables and the decomposition above, we get that the only non-zero terms of the type $\langle X^2 Y \rangle_0$ with $X, Y \in \{A, B_H, B_F, B_X\}$ yield the contribution:

$$\begin{aligned} &3\langle A^2 (B_H + B_F + B_X) \rangle_0 = \\ &\frac{3}{n^{3/2}} \sum_{\mu,i,l} [g_{\mu l} r_{\mu l} \kappa^{(3,F)}_{\mu i} \kappa^{(3,X)}_{il} + \kappa^{(3,H)}_{\mu l} m^F_{\mu i} v^F_{\mu i} \kappa^{(3,X)}_{il} + \kappa^{(3,H)}_{\mu l} \kappa^{(3,F)}_{\mu i} m^X_{il} v^X_{il}]. \end{aligned} \tag{A.12}$$

Finally, the last contribution to $\langle U^3 \rangle_0$ comes from the term:

$$\begin{aligned} &6\langle A B_H B_F + A B_H B_X + A B_F B_X + B_H B_F B_X \rangle_0 = \\ &\frac{6}{n^{3/2}} \sum_{\mu,i,l} [g_{\mu l} r_{\mu l} m^F_{\mu i} v^F_{\mu i} \kappa^{(3,X)}_{il} + g_{\mu l} r_{\mu l} \kappa^{(3,F)}_{\mu i} m^X_{il} v^X_{il} + \kappa^{(3,H)}_{\mu l} m^F_{\mu i} v^F_{\mu i} m^X_{il} v^X_{il} + g_{\mu l} r_{\mu l} m^F_{\mu i} v^F_{\mu i} m^X_{il} v^X_{il}]. \end{aligned} \tag{A.13}$$

Summing the contributions from eqs. (A.11),(A.12),(A.13) yields $\langle U^3 \rangle_0$, which gives:

$$\begin{aligned} &\frac{1}{3!} \Big(\frac{\partial^3 \Phi_{\mathbf{Y},n}}{\partial \eta^3} \Big)_{\eta=0} = \frac{-1}{6n^{5/2}(m+p)} \sum_{\mu,i,l} [\kappa^{(3,H)}_{\mu l} \kappa^{(3,F)}_{\mu i} \kappa^{(3,X)}_{il} - g^3_{\mu l} \kappa^{(3,F)}_{\mu i} \kappa^{(3,X)}_{il} + \kappa^{(3,H)}_{\mu l} \kappa^{(3,F)}_{\mu i} (m^X_{il})^2 \\ &+ \kappa^{(3,H)}_{\mu l} (m^F_{\mu i})^3 \kappa^{(3,X)}_{il} + 3 g_{\mu l} r_{\mu l} \kappa^{(3,F)}_{\mu i} \kappa^{(3,X)}_{il} + 3 \kappa^{(3,H)}_{\mu l} m^F_{\mu i} v^F_{\mu i} \kappa^{(3,X)}_{il} + 3 \kappa^{(3,H)}_{\mu l} \kappa^{(3,X)}_{\mu i} m^X_{il} v^X_{il} \\ &+ 6 g_{\mu l} r_{\mu l} m^F_{\mu i} v^F_{\mu i} \kappa^{(3,X)}_{il} + 6 g_{\mu l} r_{\mu l} \kappa^{(3,F)}_{\mu i} m^X_{il} v^X_{il} + 6 \kappa^{(3,H)}_{\mu l} m^F_{\mu i} v^F_{\mu i} m^X_{il} v^X_{il} + 6 g_{\mu l} r_{\mu l} m^F_{\mu i} v^F_{\mu i} m^X_{il} v^X_{il}]. \end{aligned}$$

Since all involved terms inside the sum are of order $\mathcal{O}_n(1)$, and given the global scaling factor, it is clear that the third order is subdominant:

$$\frac{1}{3!} \Big(\frac{\partial^3 \Phi_n}{\partial \eta^3} \Big)_{\eta=0} = o_n(1). \tag{A.14}$$

On higher order moments – An important remark that one can already conjecture by generalizing from these arguments is that at any given order of perturbation in η, *only the first two moments of the fields* $\boldsymbol{H}, \boldsymbol{F}, \boldsymbol{X}$ *will appear at dominant order*. This conjecture arises as a consequence of a scaling argument: in the terms remaining in the Plefka expansion, the higher-order moments constraint too many indices, and thus the terms involving them can be neglected. Note that we showed this property for a large class of inference models in Chapter 2.

A.3.3 Order 4 in η

This section describes the calculation of the order 4 perturbation of the free entropy. It is particularly tedious and lengthy, but the techniques involved are not conceptually complicated. We start with eq. (2.7d). For simplicity, we will not consider terms involving cumulants of order

3 and 4 of the variables $ih_{\mu l}$, $f_{\mu i}$ and x_{il}. One can check that the combination of the terms containing these moments would not give a thermodynamically relevant contribution. From the equations (3.12) at order 2 and the Maxwell relations (e.g. eq. (A.7)) we obtain the derivatives of the Lagrange multipliers, at leading order and at $\eta = 0$:

$$\begin{cases} \partial_\eta^2 \omega_{\mu l} &= \frac{2}{n} g_{\mu l} \sum_i [(m^F_{\mu i})^2 v^X_{il} + v^F_{\mu i}(m^X_{il})^2], \\ \partial_\eta^2 b_{\mu l} &= -\frac{2}{n} \sum_i [v^F_{\mu i} v^X_{il} + (m^F_{\mu i})^2 v^X_{il} + v^F_{\mu i}(m^X_{il})^2], \\ \partial_\eta^2 \lambda^F_{\mu i} &= \frac{2}{n} m^F_{\mu i} \sum_l [r_{\mu l}(m^X_{il})^2 - g^2_{\mu l} v^X_{il}], \\ \partial_\eta^2 \gamma^F_{\mu i} &= -\frac{2}{n} \sum_l [r_{\mu l} v^X_{il} + r_{\mu l}(m^X_{il})^2 - g^2_{\mu l} v^X_{il}], \\ \partial_\eta^2 \lambda^X_{il} &= \frac{2}{n} m^X_{il} \sum_\mu [r_{\mu l}(m^F_{\mu i})^2 - g^2_{\mu l} v^F_{\mu i}], \\ \partial_\eta^2 \gamma^X_{il} &= -\frac{2}{n} \sum_\mu [r_{\mu l} v^F_{\mu i} + r_{\mu l}(m^F_{\mu i})^2 - g^2_{\mu l} v^F_{\mu i}]. \end{cases} \tag{A.15}$$

We can compute the following averages:

$$\langle U^2 x_{il} \rangle_0 = -\frac{2}{n} m^X_{il} v^X_{il} \sum_\mu r_{\mu l} v^F_{\mu i},$$

$$\langle U^2 (x^2_{il} + 2m^X_{il} x_{il} - v^X_{il}) \rangle_0 = \frac{2(v^X_{il})^2}{n} \sum_\mu [-r_{\mu l} v^F_{\mu i} + g^2_{\mu l} v^F_{\mu i} - r_{\mu l}(m^F_{\mu i})^2] - \frac{4}{n} v^X_{il}(m^X_{il})^2 \sum_\mu r_{\mu l} v^F_{\mu i}.$$

From this and eq. (A.15), one can obtain the term involving the derivatives of the Lagrange parameters λ^X and γ^X in eq. (2.7d):

$$-3 \sum_{i,l} \Big[\partial_\eta^2 \lambda^X_{il} \langle U^2 x_{il} \rangle_0 + \frac{\partial_\eta^2 \gamma^X_{il}}{2} \langle U^2 (x^2_{il} + 2m^X_{il} x_{il} - v^X_{il}) \rangle_0 \Big] = \tag{A.16}$$
$$- \frac{12}{n^2} \sum_{i,l} v^X_{il} (m^X_{il})^2 \Big[\sum_\mu r_{\mu l} v^F_{\mu i} \Big]^2 - \frac{6}{n^2} \sum_{i,l} (v^X_{il})^2 \Big[\sum_\mu [r_{\mu l} v^F_{\mu i} + r_{\mu l}(m^F_{\mu i})^2 - g^2_{\mu l} v^F_{\mu i}] \Big]^2.$$

Doing similarly for $(ih)_{\mu l}$ and $f_{\mu i}$, we obtain all the terms involving derivatives of the Lagrange multipliers:

$$-3 \sum_{\mu,i} \Big[\partial_\eta^2 \lambda^F_{\mu i} \langle U^2 f_{\mu i} \rangle_0 + \frac{\partial_\eta^2 \gamma^F_{\mu i}}{2} \langle U^2 (f^2_{\mu i} + 2m^F_{\mu i} f_{\mu i} - v^F_{\mu i}) \rangle_0 \Big] = \tag{A.17}$$
$$- \frac{12}{n^2} \sum_{\mu,i} v^F_{\mu i} (m^F_{\mu i})^2 \Big[\sum_l r_{\mu l} v^X_{il} \Big]^2 - \frac{6}{n^2} \sum_{\mu,i} (v^F_{\mu i})^2 \Big[\sum_l [r_{\mu l} v^X_{il} + r_{\mu l}(m^X_{il})^2 - g^2_{\mu l} v^X_{il}] \Big]^2,$$

$$-3 \sum_{\mu,l} \Big[\partial_\eta^2 \omega_{\mu l} \langle U^2 (ih)_{\mu l} \rangle_0 + \frac{\partial_\eta^2 b_{\mu l}}{2} \langle U^2 ((ih)^2_{\mu l} - 2g_{\mu l}(ih)_{\mu l} + r_{\mu l}) \rangle_0 \Big] = \tag{A.18}$$
$$\frac{12}{n^2} \sum_{\mu,l} r_{\mu l} g^2_{\mu l} \Big[\sum_i v^F_{\mu i} v^X_{il} \Big]^2 - \frac{6}{n^2} \sum_{\mu,l} r^2_{\mu l} \Big[\sum_i [v^F_{\mu i} v^X_{il} + (m^F_{\mu i})^2 v^X_{il} + v^F_{\mu i}(m^X_{il})^2] \Big]^2.$$

The calculation at order 2 already gave:

$$-3 \langle U^2 \rangle_0^2 = -\frac{3}{n^2} \sum_{\substack{\mu,i,l \\ \mu' i' l'}} [-r_{\mu l} v^F_{\mu i} v^X_{il} + g^2_{\mu l} v^F_{\mu i} v^X_{il} - r_{\mu l}(m^F_{\mu i})^2 v^X_{il} - r_{\mu l} v^F_{\mu i}(m^X_{il})^2] \tag{A.19}$$
$$\times [-r_{\mu' l'} v^F_{\mu' i'} v^X_{i' l'} + g^2_{\mu' l'} v^F_{\mu' i'} v^X_{i' l'} - r_{\mu' l'}(m^F_{\mu' i'})^2 v^X_{i' l'} - r_{\mu' l'} v^F_{\mu' i'}(m^X_{i' l'})^2].$$

We finally have to compute $\langle U^4 \rangle_0$, whose calculation is very tedious, but not conceptually difficult. We again make use of the decomposition of eq. (A.10). A first simplification arises

since the variables are centered and have negligible moments of odd order. This implies:

$$\langle U^4\rangle_0 = \underbrace{\langle A^4\rangle_0}_{I_1} + \underbrace{6\langle A^2(B_H^2+B_F^2+B_X^2)\rangle_0}_{I_2} + \underbrace{\langle (B_H+B_F+B_X)^4\rangle_0}_{I_3} + \mathcal{O}_n(1).$$

We now compute the three terms I_1, I_2, I_3 independently. An explicit calculation gives

$$\begin{aligned} I_1 &= \frac{6}{n^2}\sum_{\mu,i,l} r_{\mu l}^2 (v_{\mu i}^F)^2 (v_{il}^X)^2 + \frac{3}{n^2}\sum_{\substack{\mu,i,l\\ \mu' i' l'}} r_{\mu l} r_{\mu' l'} v_{\mu i}^F v_{\mu' i'}^F v_{il}^X v_{i' l'}^X \\ &+ \frac{6}{n^2}\sum_{\mu,i,l} r_{\mu l} v_{\mu i}^F v_{il}^X \Big(\sum_{\mu'} r_{\mu' l} v_{\mu' i}^F v_{il}^X + \sum_{i'} r_{\mu l} v_{\mu i'}^F v_{i' l}^X + \sum_{l'} r_{\mu l'} v_{\mu i}^F v_{i l'}^X \Big) + \mathcal{O}_n(1). \end{aligned} \tag{A.20}$$

Importantly, here the indices are not supposed to be pairwise distinct unless explicitly stated so. The terms I_1, I_2 can be explicitly computed as well, and are very lengthy:

$$\begin{aligned} I_2 &= \frac{6}{n^2}\sum_{\substack{\mu,i,l\\ \mu' i' l'}} r_{\mu l} v_{\mu i}^F v_{il}^X [-g_{\mu' l'}^2 v_{\mu' i'}^F v_{i' l'}^X + r_{\mu' l'} v_{\mu' i'}^F (m_{i' l'}^X)^2 + r_{\mu' l'} (m_{\mu' i'}^F)^2 v_{il}^X] \\ &+ \frac{12}{n^2}\sum_{\mu,i,l} r_{\mu l} v_{\mu i}^F v_{il}^X [-g_{\mu l}^2 v_{\mu i}^F v_{il}^X + r_{\mu l} v_{\mu i}^F (m_{il}^X)^2 + r_{\mu l} (m_{\mu i}^F)^2 v_{il}^X] \\ &+ \frac{12}{n^2}\sum_{\mu,i,l}\sum_{\mu'} r_{\mu l} v_{\mu i}^F v_{il}^X [-g_{\mu' l}^2 v_{\mu' i}^F v_{il}^X + r_{\mu' l} v_{\mu' i}^F (m_{il}^X)^2 + r_{\mu' l} (m_{\mu' i}^F)^2 v_{il}^X] \\ &+ \frac{12}{n^2}\sum_{\mu,i,l}\sum_{i'} r_{\mu l} v_{\mu i}^F v_{il}^X [-g_{\mu l}^2 v_{\mu i'}^F v_{i' l}^X + r_{\mu l} v_{\mu i'}^F (m_{i' l}^X)^2 + r_{\mu l} (m_{\mu i'}^F)^2 v_{i' l}^X] \\ &+ \frac{12}{n^2}\sum_{\mu,i,l}\sum_{l'} r_{\mu l} v_{\mu i}^F v_{il}^X [-g_{\mu l'}^2 v_{\mu i}^F v_{i l'}^X + r_{\mu l'} v_{\mu i}^F (m_{il}^X)^2 + r_{\mu l'} (m_{\mu i}^F)^2 v_{i l'}^X] + \mathcal{O}_n(1). \end{aligned} \tag{A.21}$$

$$\begin{aligned} I_3 &= \frac{3}{n^2}\sum_{\substack{\mu,i,l\\ \mu' i' l'}} [g_{\mu l}^2 g_{\mu' l'}^2 v_{\mu i}^F v_{\mu' i'}^F v_{il}^X v_{i' l'}^X + r_{\mu l} r_{\mu' l'} (m_{\mu i}^F)^2 (m_{\mu' i'}^F)^2 v_{il}^X v_{i' l'}^X + r_{\mu l} r_{\mu' l'} v_{\mu i}^F v_{\mu' i'}^F (m_{il}^X)^2 (m_{i' l'}^X)^2] \\ &- \frac{6}{n^2}\sum_{\substack{\mu,i,l\\ \mu' i' l'}} [-r_{\mu l} r_{\mu' l'} (m_{\mu' i'}^F)^2 v_{\mu i}^F (m_{il}^X)^2 v_{i' l'}^X + g_{\mu l}^2 r_{\mu' l'} (m_{\mu' i'}^F)^2 v_{\mu i}^F v_{il}^X v_{i' l'}^X + g_{\mu l}^2 r_{\mu' l'} v_{\mu' i'}^F v_{\mu i}^F (m_{i' l'}^X)^2 v_{il}^X] \\ &- \frac{12}{n^2}\sum_{\mu,i,l} \Big[\sum_{\mu'} g_{\mu l}^2 r_{\mu' l} (m_{\mu i}^F)^2 v_{\mu i}^F (v_{il}^X)^2 - \sum_{i'} r_{\mu l}^2 (m_{\mu i'}^F)^2 v_{\mu i}^F (m_{il}^X)^2 v_{i' l}^X + \sum_{l'} g_{\mu l}^2 r_{\mu l'} (v_{\mu i}^F)^2 (m_{i l'}^X)^2 v_{il}^X \Big] \\ &+ \frac{6}{n^2}\sum_{\mu,i,l} \Big[\sum_{\mu',i'} m_{\mu i}^F m_{\mu' i}^F m_{\mu' i'}^F m_{\mu i'}^F r_{\mu l} r_{\mu' l} v_{il}^X v_{i' l}^X + \sum_{i',l'} m_{il}^X m_{i' l}^X m_{i' l'}^X m_{i l'}^X v_{\mu i}^F v_{\mu i'}^F r_{\mu l} r_{\mu l'} \Big], \\ &+ \frac{6}{n^2}\sum_{\mu,i,l} \Big[\sum_{\mu',l'} g_{\mu l} g_{\mu' l} g_{\mu' l'} g_{\mu l'} v_{\mu i}^F v_{\mu' i}^F v_{il}^X v_{i l'}^X \Big] + \mathcal{O}_n(1). \end{aligned} \tag{A.22}$$

Many simplifications occur in the terms of eqs. (A.16) to (A.22). Two type of terms are for instance negligible:

- Terms of the type $n^{-4}\sum_{\mu i l} A_{\mu i l}$, with $A_{\mu i l}$ typically of order 1. These terms are negligible by a simple scaling argument.

- Terms involving strongly irreducible diagrams of m^F or m^X. For instance, the term:

$$\sum_{\mu}\sum_{i\neq i'}\sum_{l\neq l'} m^X_{il}\, m^X_{il'} m^X_{i'l'} m^X_{i'l} r_{\mu l} r_{\mu l'} v^F_{\mu i} v^F_{\mu i'}.$$

  By H.1, the variables $\mathbf{m}^F, \mathbf{m}^X$ behave like uncorrelated variables, so that all these diagrams (including the simple loops) will be negligible. This is precisely the sort of terms that is not negligible when involving $g_{\mu l}$, as a consequence of H.2.

We can now sum all eqs. (A.16) to (A.22), simplifying the terms that are negligible by the arguments above, and checking that almost all non-negligible terms are cancelling each other. This is a lengthy but straightforwards calculation, and we reach:

$$\frac{1}{4!}\Big(\frac{\partial^4 \Phi_{\mathbf{Y},n}}{\partial \eta^4}\Big)_0 = \frac{1}{4n^3(m+p)} \sum_i \sum_{\mu_1\neq\mu_2} \sum_{l_1\neq l_2} g_{\mu_1 l_1} g_{\mu_1 l_2} g_{\mu_2 l_2} g_{\mu_2 l_1} v^F_{\mu_1 i} v^F_{\mu_2 i} v^X_{i l_1} v^X_{i l_2} + \mathcal{O}_n(1).$$

A.4 The expansion for symmetric extensive-rank matrix factorization

For the symmetric Model $\mathbf{XX}^\intercal$, we can perform a PGY expansion in a very similar way to the case of Model $\mathbf{FX}$ that we just described in Section A.3. The majority of the calculation is straightforwardly transposed, and we briefly outline its main steps here. Recall the free entropy of eq. (3.3b):

$$\Phi_{\mathbf{Y},n} \equiv \frac{1}{nm} \ln \int \prod_{\mu,i} P_X(\mathrm{d}X_{\mu i}) \prod_{\mu<\nu} P_{\text{out}}\Big(Y_{\mu\nu}\Big|\frac{1}{\sqrt{n}}\sum_i X_{\mu i} X_{\nu i}\Big).$$

Introducing a field $\mathbf{h} \equiv \mathbf{XX}^\intercal/\sqrt{n}$, its Fourier conjugate $\mathbf{H}$, and Lagrange multipliers to fix the first and second moments of both $\mathbf{X}$ and $\mathbf{H}$, we reach a very similar form to eq. (3.8):

$$\begin{aligned} nm\Phi_{\mathbf{Y},n} = &\sum_{\mu,i}\Big[\lambda_{\mu i} m_{\mu i} + \frac{\gamma_{\mu i}}{2}(v_{\mu i} + (m_{\mu i})^2)\Big] + \sum_{\mu<\nu}\Big[-\omega_{\mu\nu} g_{\mu\nu} - \frac{b_{\mu\nu}}{2}(-r_{\mu\nu} + g^2_{\mu\nu})\Big] \\ &+ \ln \int P_H(\mathrm{d}\mathbf{H})\, P_X(\mathrm{d}\mathbf{X})\, e^{-S_{\text{eff}}[\mathbf{X},\mathbf{H}]}, \end{aligned} \tag{A.23}$$

in which we introduced the *effective action*:

$$\begin{aligned} S_{\text{eff}}[\mathbf{X},\mathbf{H}] = &\sum_{\mu,i}\Big[\lambda_{\mu i} X_{\mu i} + \frac{\gamma_{\mu i}}{2} X^2_{\mu i}\Big] + \sum_{\mu<\nu}\Big[\omega_{\mu\nu}(iH)_{\mu\nu} - \frac{b_{\mu\nu}}{2}(iH)^2_{\mu\nu}\Big] \\ &+ \underbrace{\frac{1}{\sqrt{n}}\sum_{\mu<\nu}\sum_i (iH)_{\mu\nu} X_{\mu i} X_{\nu i}}_{H_{\text{eff}}}. \end{aligned} \tag{A.24}$$

The (un-normalized) distribution over $H_{\mu\nu}$ is given by the Fourier transform of the channel:

$$P^{\mu\nu}_H[\mathrm{d}H] \equiv \int \frac{\mathrm{d}\tilde{H}}{2\pi} e^{iH\tilde{H}} P_{\text{out}}(Y_{\mu\nu}|\tilde{H}).$$

Again, we introduce a factor η in front of H_{eff} in eq. (A.24), and we expand the free entropy as a function of η. Computing the first order of perturbation and the operator U of Georges-Yedidia

is readily done exactly as in Section A.3.1, and we reach:

$$\Big(\frac{\partial \Phi_{\mathbf{Y},n}}{\partial \eta}\Big)_{\eta=0} = \frac{1}{n^{3/2}m}\sum_i \sum_{\mu<\nu} g_{\mu\nu} m_{\mu i} m_{\nu i}, \tag{A.25}$$
$$U_{\mathbf{Y},\eta=0} = \frac{1}{\sqrt{n}}\sum_i \sum_{\mu<\nu} [(ih)_{\mu\nu} x_{\mu i} x_{\nu i} - g_{\mu\nu} x_{\mu i} x_{\nu i} + (ih)_{\mu\nu} m_{\mu i} x_{\nu i} + (ih)_{\mu\nu} x_{\mu i} m_{\nu i}],$$

in which the lowercase variables again designate the centered variables. From the expression of U we can obtain the following orders of perturbation:

$$\begin{cases} \frac{1}{2}\Big(\frac{\partial^2 \Phi_{\mathbf{Y},n}}{\partial \eta^2}\Big)_{\eta=0} &= \frac{1}{2n^2 m}\sum_i \sum_{\mu<\nu} [-r_{\mu\nu} v_{\mu i} v_{\nu i} + g_{\mu\nu}^2 v_{\mu i} v_{\nu i} - r_{\mu\nu} m_{\mu i}^2 v_{\nu i} - r_{\mu\nu} v_{\mu i} m_{\nu i}^2], \\ \frac{1}{3!}\Big(\frac{\partial^3 \Phi_{\mathbf{Y},n}}{\partial \eta^3}\Big)_{\eta=0} &= \frac{1}{6n^{5/2}m}\sum_i \sum_{\substack{\mu_1,\mu_2,\mu_3 \\ \text{pairwise distinct}}} g_{\mu_1\mu_2} g_{\mu_2\mu_3} g_{\mu_3\mu_1} \prod_{a=1}^{3} v_{\mu_a i} + \mathcal{O}_n(1), \\ \frac{1}{4!}\Big(\frac{\partial^4 \Phi_{\mathbf{Y},n}}{\partial \eta^4}\Big)_{\eta=0} &= \frac{1}{8n^3 m}\sum_i \sum_{\substack{\mu_1,\mu_2,\mu_3,\mu_4 \\ \text{pairwise distinct}}} g_{\mu_1\mu_2} g_{\mu_2\mu_3} g_{\mu_3\mu_4} g_{\mu_4\mu_1} \prod_{a=1}^{4} v_{\mu_a i} + \mathcal{O}_n(1). \end{cases} \tag{A.26}$$

Here we symmetrized $g_{\nu\mu} \equiv g_{\mu\nu}$, and we adopted the convention $g_{\mu\mu} = 0$.

Appendix B

Details of replica computations

B.1 Replica calculation for the committee machine

Our goal here is to provide a heuristic derivation of the formula of Theorem 4.1 using the replica method, a powerful non-rigorous tool from statistical physics of disordered systems. To the best of our knowledge, this computation was first performed for a committee machine in [Sch93]. Recall that we introduced the method in Section 1.3.1, and the reader can come back to it to freshen her/his memory if necessary. As we know, in order to compute the asymptotic free entropy with the replica method we need the moments of the partition function (given by eq. (4.5)), for integer p:

$$\begin{aligned}\mathbb{E}\mathcal{Z}_n^p &= \mathbb{E}\Bigg\{\Bigg[\int_{\mathbb{R}^n\times\mathbb{R}^K} \mathrm{d}\mathbf{w}\prod_{i=1}^n P_0(\{w_{il}\}_{l=1}^K)\prod_{\mu=1}^m P_{\text{out}}\Big(Y_\mu\Big|\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{\mu i}w_{il}\Big\}_{l=1}^K\Big)\Bigg]^p\Bigg\},\\ &= \mathbb{E}\Bigg[\prod_{a=1}^p\int_{\mathbb{R}^n\times\mathbb{R}^K} \mathrm{d}\mathbf{w}^a\prod_{i=1}^n P_0(\{w_{il}^a\}_{l=1}^K)\prod_{\mu=1}^m P_{\text{out}}\Big(Y_\mu\Big|\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{\mu i}w_{il}^a\Big\}_{l=1}^K\Big)\Bigg].\end{aligned}$$

The outer expectation is done over $X_{\mu i}\sim\mathcal{N}(0,1)$, $\mathbf{W}^\star$ and $\mathbf{Y}$. Writing $\mathbf{W}^\star$ as $\mathbf{w}^0$ we have:

$$\mathbb{E}\mathcal{Z}_n^p = \mathbb{E}_{\mathbf{X}}\int_{\mathbb{R}^m}\mathrm{d}\mathbf{Y}\prod_{a=0}^p\Bigg[\int_{\mathbb{R}^n\times\mathbb{R}^K} \mathrm{d}\mathbf{w}^a\prod_{i=1}^n P_0(\{w_{il}^a\}_{l=1}^K)\prod_{\mu=1}^m P_{\text{out}}\Big(Y_\mu\Big|\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{\mu i}w_{il}^a\Big\}_{l=1}^K\Big)\Bigg].$$

To perform the average over $\mathbf{X}$ we notice that the variables $\{Z_{\mu l}^a\}$, defined by

$$Z_{\mu l}^a \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n X_{\mu i}w_{il}^a,$$

follow a multivariate Gaussian distribution with zero mean and covariance tensor:

$$\mathbb{E}Z_{\mu l}^a Z_{\nu l'}^b = \delta_{\mu\nu}\Sigma_{\underset{bl'}{al}} = \delta_{\mu\nu}Q_{bl'}^{al}, \qquad \text{with} \qquad Q_{bl'}^{al}\equiv\frac{1}{n}\sum_{i=1}^n w_{il}^a w_{il'}^b.$$

For every a,b, $Q_b^a\in\mathcal{S}_K$ is the *overlap* matrix between replicas a and b. Introducing δ functions to fix Q we arrive at :

$$\mathbb{E}[\mathcal{Z}_n^p] = \prod_{(a,r)}\int_{\mathbb{R}}\mathrm{d}Q_{ar}^{ar}\prod_{\{(a,r);(b,r')\}}\int_{\mathbb{R}}\mathrm{d}Q_{br'}^{ar}[I_{\text{prior}}(\{Q_{br'}^{ar}\})\times I_{\text{channel}}(\{Q_{br'}^{ar}\})],$$

with the two auxiliary integrals:

$$I_{\text{prior}}(\{Q^{ar}_{br'}\}) \equiv \prod_{a=0}^{p}\Big[\int \mathrm{d}\mathbf{w}^a P_0(\mathbf{w}^a)\Big]\Big[\prod_{\{(a,l);(b,l')\}}\delta\Big(Q^{al}_{bl'} - \frac{1}{n}\sum_{i=1}^{n} w^a_{il}w^b_{il'}\Big)\Big],$$

$$I_{\text{channel}}(\{Q^{ar}_{br'}\}) \equiv \int \mathrm{d}\mathbf{Y}\prod_{a=0}^{p}\int \mathrm{d}\mathbf{Z}^a \prod_{a=0}^{p} P_{\text{out}}(\mathbf{Y}|\mathbf{Z}^a)\frac{\exp\Big[-\frac{1}{2}\sum_{\mu=1}^{m}\sum_{a,b}\sum_{l,l'} Z^a_{\mu l}Z^b_{\mu l'}(\Sigma^{-1})^{al}_{bl'}\Big]}{\Big((2\pi)^{K(p+1)}\det\Sigma\Big)^{m/2}}.$$

By Fourier expanding the delta functions in I_{prior} and performing a saddle-point one obtains:

$$\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}[\mathcal{Z}^p_n] = \operatorname*{extr}_{Q,\hat{Q}}[H(Q,\hat{Q})], \tag{B.1}$$

in which (recall $m/n \to \alpha > 0$) :

$$H(Q,\hat{Q}) \equiv \frac{1}{2}\sum_{a=0}^{p}\sum_{l,l'} Q^{al}_{al}\hat{Q}^{al}_{al} - \frac{1}{2}\sum_{a\neq b}\sum_{l,l'} Q^{al}_{bl'}\hat{Q}^{al}_{bl'} + \ln I + \alpha\ln J, \tag{B.2}$$

with the two functions:

$$I \equiv \prod_{a=0}^{p}\int_{\mathbb{R}^K}\mathrm{d}w^a P_0(w^a)\exp\Big\{-\frac{1}{2}\sum_{a=0}^{p}\sum_{l,l'}\hat{Q}^{al}_{al'}w^a_l w^a_{l'} + \frac{1}{2}\sum_{a\neq b}\sum_{l,l'}\hat{Q}^{al}_{bl'}w^a_l w^b_{l'}\Big\},$$

$$J \equiv \int_{\mathbb{R}}\mathrm{d}y\prod_{a=0}^{p}\int_{\mathbb{R}^K}\frac{\mathrm{d}Z^a}{(2\pi)^{K(p+1)/2}}\frac{P_{\text{out}}(y|Z^a)}{\sqrt{\det\Sigma}}\exp\Big\{-\frac{1}{2}\sum_{a,b=0}^{p}\sum_{l,l'=1}^{K} Z^a_l Z^b_{l'}(\Sigma^{-1})^{al}_{bl'}\Big\}.$$

Our goal is to express $H(Q,\hat{Q})$ as an analytical function of p, in order to perform the replica trick. To do so, we will assume that the extremum of H is attained at a point in $Q,\hat{Q}$ space such that a *replica symmetry* property is verified. More concretely, we assume:

$$\begin{cases} \exists Q^0 \in \mathcal{S}_K \text{ s.t} \quad \forall a \in [|0,p|] \quad \forall (l,l') \in [|1,K|]^2 \quad Q^{al}_{al'} = Q^0_{ll'}, \\ \exists q \in \mathcal{S}_K \text{ s.t} \quad \forall (a<b) \in [|0,p|]^2 \quad \forall (l,l') \in [|1,K|]^2 \quad Q^{al}_{bl'} = q_{ll'}, \end{cases} \tag{B.3}$$

and similarly for $\hat{Q}$. Under the ansatz (B.3), we obtain from eq. (B.2):

$$H(Q^0,\hat{Q}^0,q,\hat{q}) = \frac{p+1}{2}\mathrm{Tr}[Q^0\hat{Q}^0] - \frac{p(p+1)}{2}\mathrm{Tr}[q\hat{q}] + \ln I + \alpha\ln J. \tag{B.4}$$

Remains now to compute an expression for I and J that is analytical in p, in order to take the limit $p \downarrow 0$. This can be done using the identity, for any $M \in \mathcal{S}^+_K$ and $x \in \mathbb{R}^K$:

$$\exp\Big(\frac{1}{2}x^\intercal M x\Big) = \int_{\mathbb{R}^K}\mathcal{D}\xi\,\exp\Big(\xi^\intercal M^{1/2}x\Big),$$

in which $\mathcal{D}\xi$ is the standard Gaussian measure on $\mathbb{R}^K$. We obtain, after a tedious algebraic calculation that is left to the reader:

$$\begin{cases} I &= \displaystyle\int_{\mathbb{R}^K}\mathcal{D}\xi\Big[\int_{\mathbb{R}^K}\mathrm{d}w\,P_0(w)\,\exp\Big[-\frac{1}{2}w^\intercal(\hat{Q}^0+\hat{q})w + \xi^\intercal\hat{q}^{1/2}w\Big]\Big]^{p+1}, \\ J &= \displaystyle\int_{\mathbb{R}}\mathrm{d}y\int_{\mathbb{R}^K}\mathcal{D}\xi\Big[\int_{\mathbb{R}^K}\mathrm{d}Z P_{\text{out}}\{y|(Q^0-q)^{1/2}Z + q^{1/2}\xi\}\Big]^{p+1}. \end{cases} \tag{B.5}$$

We now notice that the value of Q^0 is actually constrained by the problem and the replica symmetry. Indeed, our calculation must be consistent in the sense that if taking $p = 0$ in eq. (B.1), we must find $\mathrm{extr}_{Q,\hat{Q}}[H(Q,\hat{Q})] = 0$ at $p = 0$.

In the $p \downarrow 0$ limit, one easily gets $J = 1$ and $I = \int_{\mathbb{R}^K} \mathrm{d}w\, P_0(w) \exp[-w^\intercal \hat{Q}^0 w^0/2]$. Extremizing over $Q^0, \hat{Q}^0$ implies that they satisfy $\hat{Q}^0 = 0$ and $Q^0 = \rho$ (the covariance matrix of P_0). We can now plug in these values in eq. (B.5), and then take the $p \downarrow 0$ limit of $H(q,\hat{q})/p$ in eq. (B.4). In the end, by eq. (B.1) and the famous "replica trick" we obtain the final formula for the free entropy:

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_n = \operatorname*{extr}_{q,\hat{q}} \{ -\frac{1}{2}\mathrm{Tr}[q\hat{q}] + I_P + \alpha I_C \}, \tag{B.6}$$

wih the two functions:

$$I_P \equiv \int_{\mathbb{R}^K} \mathcal{D}\xi \mathrm{d}w^0 P_0(w^0) e^{-\frac{1}{2}(w^0)^\intercal \hat{q} w^0 + \xi^\intercal \hat{q}^{1/2} w^0} \ln \int_{\mathbb{R}^K} \mathrm{d}w P_0(w) e^{-\frac{1}{2} w^\intercal \hat{q} w + \xi^\intercal \hat{q}^{1/2} w},$$
$$I_C \equiv \int \mathrm{d}y \int_{\mathbb{R}^K} \mathcal{D}\xi\, \mathcal{D}Z^0\, P_{\mathrm{out}}\{y|(Q^0 - q)^{1/2} Z^0 + q^{1/2}\xi\} \ln \int_{\mathbb{R}^K} \mathcal{D}Z\, P_{\mathrm{out}}\{y|(Q^0 - q)^{1/2} Z + q^{1/2}\xi\}.$$

A known ambiguity of the replica method is that its result is given as an extremum, here over the set $\mathcal{S}_K^+(\rho)$. Assuming that this extr is realized as a $\sup_{\hat{q}} \inf_q$, one can easily see that eq. (B.6) yields back the claim of Theorem 4.1. Our rigorous proof, sketched in Section 4.4, will allow to lift the ambiguity of the extremum.

B.2 Replica computation for generic GLMs

In this section, we perform the step-by-step replica calculation that gives Conjecture 6.1.

B.2.1 Setting and strategy

We let $n, m \to \infty$ with $m/n \to \alpha > 0$. Recall that we are interested in the partition function:

$$\mathcal{Z}_n(\mathbf{Y}) \equiv \int_{\mathbb{K}^n} \prod_{i=1}^n P_0(\mathrm{d}x_i) \prod_{\mu=1}^m P_{\mathrm{out}}\Big(Y_\mu \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_{\mu i} x_i\Big).$$

Here $\mathbf{\Phi}$ is a matrix that is left and right-orthogonally (unitarily) invariant, meaning that for all $\mathbf{O}, \mathbf{U} \in \mathcal{U}_\beta(m) \times \mathcal{U}_\beta(n)$, $\mathbf{\Phi} \overset{\mathrm{d}}{=} \mathbf{O}\mathbf{\Phi}\mathbf{U}$. Compared to Conjecture 6.1, we added a left-invariance hypothesis. However the analysis of G-VAMP [SRF16, RSF17] shows that this left invariance is actually not needed for the result, and thus we state Conjecture 6.1 for matrices that are only right-invariant, but we use the left invariance to simplify the following (heuristic) calculation. Moreover, we assume that the LSD ν of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ is well-defined, and that the eigenvalue distribution of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ has large deviations in a scale at least $n^{1+\eta}$ for an $\eta > 0$.

Recall that the replica trick introduced in Section 1.3.1 consists in computing the p-th moment of the partition function for arbitrary integer p, before extending this expression analytically to any $p > 0$ and using the *replica trick*:

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\mathbf{\Phi},\mathbf{Y}} \ln \mathcal{Z}_n(\mathbf{Y}) = \lim_{p\downarrow 0} \lim_{n\to\infty} \frac{1}{np} \ln \mathbb{E}_{\mathbf{\Phi},\mathbf{Y}}[\mathcal{Z}_n(\mathbf{Y})^p].$$

This method is obviously heuristic given the inversion of limits $p \downarrow 0$ and $n \to \infty$, as well as the analytic continuation to arbitrary $p > 0$ of the p-th moment.

B.2.2 Computing the p-th moment of the partition function

Thanks to Bayes-optimality, we can easily write the average of $\mathcal{Z}_n(\mathbf{Y})^p$ as an average over $p+1$ replicas of the system, by considering $\mathbf{X}^\star$ as the replica of index 0. We obtain for any $p \geq 1$:

$$\mathbb{E}[\mathcal{Z}_n(\mathbf{Y})^p] = \mathbb{E}_{\mathbf{\Phi}} \int_{\mathbb{R}^m} \mathrm{d}\mathbf{Y} \prod_{a=0}^{p} \Big\{ \Big[\int_{\mathbb{K}} \prod_{i=1}^{n} P_0(\mathrm{d}x_i^a) \int_{\mathbb{K}} \prod_{\mu=1}^{m} \mathrm{d}z_\mu^a P_{\text{out}}(Y_\mu | z_\mu^a) \Big] \delta\Big(\mathbf{z}^a - \frac{\mathbf{\Phi}\mathbf{x}^a}{\sqrt{n}}\Big) \Big\}. \tag{B.7}$$

The first step is to decompose eq. (B.7) into three terms, corresponding to the prior P_0, the channel P_{out}, and the "delta" term. Note that the matrix $\mathbf{\Phi}$ only appears in the last "delta" term. By left and right invariance of $\mathbf{\Phi}$, the quantity

$$\mathbb{E}_{\mathbf{\Phi}} \Big[\prod_{a=0}^{p} \delta\Big(\mathbf{z}^a - \frac{1}{\sqrt{n}} \mathbf{\Phi}\mathbf{x}^a\Big) \Big]$$

is completely determined by the *overlaps* $\mathbf{Q}^z \equiv \{(\mathbf{z}^a)^\dagger \mathbf{z}^b / m\}_{a,b=0}^p$ and $\mathbf{Q}^x \equiv \{(\mathbf{x}^a)^\dagger \mathbf{x}^b / n\}_{a,b=0}^p$, which are positive Hermitian matrices. As is standard in such replica calculations, we will constraint the terms in eq. (B.7) by the value of these overlaps, before performing a Laplace method on the resulting function of the overlaps. By $A_n \simeq B_n$, we will mean $(\ln A_n)/n = (\ln B_n)/n + \mathcal{O}_n(1)$. We introduce in eq. (B.7) the term:

$$1 \simeq \int \prod_{0 \leq a \leq b \leq p} \mathrm{d}Q_{ab}^x \, \mathrm{d}Q_{ab}^z \Big[\prod_{a \leq b} \delta(n Q_{ab}^x - (\mathbf{x}^a)^\dagger \mathbf{x}^b) \Big] \Big[\prod_{a \leq b} \delta(m Q_{ab}^z - (\mathbf{z}^a)^\dagger \mathbf{z}^b) \Big].$$

We can use a Fourier transformation of the delta terms, which allows in the end to transform eq. (B.7) into the product of three independent terms. Performing then Laplace's method on $\mathbf{Q}^x, \mathbf{Q}^z$ we obtain:

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}_{\mathbf{Y},\mathbf{\Phi}}[\mathcal{Z}_n(\mathbf{Y})^p] = \sup_{\mathbf{Q}^x, \mathbf{Q}^z} [I_0(p, \mathbf{Q}^x) + \alpha I_{\text{out}}(p, \mathbf{Q}^z) + I_{\text{int}}(p, \mathbf{Q}^x, \mathbf{Q}^z)],$$

in which the supremum is made over positive symmetric (Hermitian) matrices, and I_0, I_{out} and I_{int} are functions whose calculation will be detailed below.

The prior term $I_0(p, \mathbf{Q}^x)$

We have after Fourier transformation of the delta terms:

$$\begin{aligned} I_0(p, \mathbf{Q}^x) &\simeq \frac{1}{n} \ln \int \prod_{0 \leq a \leq b \leq p} \mathrm{d}\hat{Q}_{ab}^x \int_{\mathbb{K}} \prod_{a=0}^{p} \prod_{i=1}^{n} P_0(\mathrm{d}x_i^a) e^{-\frac{\beta}{2} \sum_{a,b=0}^{p} \hat{Q}_{ab}^x (\sum_i \overline{x_i^a} x_i^b - n Q_{ab}^x)}, \\ &\simeq \inf_{\hat{\mathbf{Q}}^x} \Big[\frac{\beta}{2} \sum_{a,b} Q_{ab}^x \hat{Q}_{ab}^x + \ln \int_{\mathbb{K}} \prod_{a=0}^{p} P_0(\mathrm{d}x^a) e^{-\frac{\beta}{2} \sum_{a,b} \hat{Q}_{ab}^x \overline{x^a} x^b} \Big]. \end{aligned}$$

The infimum is again over positive symmetric (Hermitian) matrices. We also made use of the fact that the prior P_0 is i.i.d. over the elements of $\mathbf{x}$.

A very important assumption that one can then use is *replica symmetry*: it amounts to assume that all the $(p+1)$ replicas are equivalent, and that this symmetry is not broken by the system at the solution of the variational principle on $\mathbf{Q}, \hat{\mathbf{Q}}$. It has been shown that for an inference problem in the Bayes-optimal setting (as is the present case), replica symmetry is never broken [ZK16]. For more details on replica symmetry, the reader can refer to Section 1.3.1. We therefore

assume a replica symmetric form of $\mathbf{Q}^x, \hat{\mathbf{Q}}^x$ at the point at which the extremum is reached:

$$\mathbf{Q}^x = \begin{pmatrix} Q_x & q_x & \cdots & q_x \\ q_x & Q_x & \cdots & q_x \\ \vdots & \vdots & \ddots & \vdots \\ q_x & q_x & \cdots & Q_x \end{pmatrix}, \qquad \hat{\mathbf{Q}}^x = \begin{pmatrix} \hat{Q}_x & -\hat{q}_x & \cdots & -\hat{q}_x \\ -\hat{q}_x & \hat{Q}_x & \cdots & -\hat{q}_x \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{q}_x & -\hat{q}_x & \cdots & \hat{Q}_x \end{pmatrix}. \tag{B.8}$$

Note that we have $Q_x, q_x, \hat{Q}_x, \hat{q}_x \in \mathbb{R}$. After a simple Gaussian transformation of the squared term using the general identity $\exp(\beta|x|^2/2) = \int_{\mathbb{K}} \mathcal{D}_\beta \xi \, \exp(\beta x \cdot \xi)$ we reach the final expression:

$$I_0(p, Q_x, q_x) = \tag{B.9}$$
$$\inf_{\hat{Q}_x, \hat{q}_x} \Big\{ \frac{\beta(p+1)}{2} Q_x \hat{Q}_x - \frac{\beta p(p+1)}{2} q_x \hat{q}_x + \ln \int_{\mathbb{K}} \mathcal{D}_\beta \xi \Big[\int_{\mathbb{K}} P_0(\mathrm{d}x) e^{-\frac{\beta(\hat{Q}_x + \hat{q}_x)}{2}|x|^2 + \beta\sqrt{\hat{q}_x} x \cdot \xi} \Big]^{p+1} \Big\}.$$

The channel term $I_{\text{out}}(p, \mathbf{Q}^z)$

This term is very similar to the prior term we just detailed. We use completely similar replica symmetric assumptions for the overlaps $\mathbf{Q}^z$ to the ones on $\mathbf{Q}^x$ described in eq. (B.8). We reach:

$$I_{\text{out}}(p, Q_z, q_z) = \inf_{\hat{Q}_z, \hat{q}_z} \Big\{ \frac{\beta(p+1)}{2} Q_z \hat{Q}_z - \frac{\beta p(p+1)}{2} q_z \hat{q}_z + \frac{\beta(p+1)}{2} \ln(2\pi/(\beta \hat{Q}_z)) \tag{B.10}$$
$$+ \ln \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{K}} \mathcal{D}_\beta \xi \Big[\int_{\mathbb{K}} \mathrm{d}z \Big(\frac{2\pi}{\beta \hat{Q}_z} \Big)^{-\beta/2} P_{\text{out}}(y|z) \; e^{-\beta \frac{\hat{Q}_z + \hat{q}_z}{2}|z|^2 + \beta\sqrt{\hat{q}_z} z \cdot \xi} \Big]^{p+1} \Big\}.$$

We normalized the integrals so that in the limit $p \downarrow 0$, the term inside the logarithm goes to 1, which will be a useful remark.

The "delta" term $I_{\text{int}}(p, \mathbf{Q}^x, \mathbf{Q}^z)$

We now turn to the computation of the "delta" term:

$$I_{\text{int}}(p, \mathbf{Q}^x, \mathbf{Q}^z) \equiv \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}_{\mathbf{\Phi}} \Big[\prod_{a=0}^{p} \delta\Big(\mathbf{z}^a - \frac{1}{\sqrt{n}} \mathbf{\Phi} \mathbf{x}^a \Big) \Big], \tag{B.11}$$

assuming that $\mathbf{Q}^x, \mathbf{Q}^z$ are known. Computing this term is central in this replica calculation. We use, as is done in [TK20], the identity:

$$\frac{1}{n} \ln \mathbb{E}_{\mathbf{\Phi}} \Big[\prod_{a=0}^{p} \delta\Big(\mathbf{z}^a - \frac{1}{\sqrt{n}} \mathbf{\Phi} \mathbf{x}^a \Big) \Big] = \lim_{\epsilon \downarrow 0} \frac{1}{n} \ln \mathbb{E}_{\mathbf{\Phi}} \Big[\frac{\exp\big\{ -\frac{\beta}{2\epsilon} \sum_a \big\| \mathbf{z}^a - \frac{1}{\sqrt{n}} \mathbf{\Phi} \mathbf{x}^a \big\|^2 \big\}}{(2\pi\epsilon/\beta)^{\frac{\beta m(p+1)}{2}}} \Big], \tag{B.12}$$

and we invert the $n \to \infty$ and the $\epsilon \downarrow 0$ limit. Let us rewrite the RHS of eq. (B.12). As $\mathbf{\Phi}$ is orthogonally (unitarily) invariant, we can write it as:

$$\mathbb{E}_{\mathbf{\Phi}} \Big[\frac{\exp\big\{ -\frac{\beta}{2\epsilon} \sum_a \big\| \mathbf{z}^a - \frac{1}{\sqrt{n}} \mathbf{\Phi} \mathbf{x}^a \big\|^2 \big\}}{(2\pi\epsilon/\beta)^{\frac{\beta m(p+1)}{2}}} \Big] = \mathbb{E}_{\mathbf{\Phi}, \mathbf{O}, \mathbf{U}} \Big[\frac{\exp\Big\{ -\frac{\beta}{2\epsilon} \sum_a \big\| \mathbf{O} \mathbf{z}^a - \frac{1}{\sqrt{n}} \mathbf{\Phi} \mathbf{U} \mathbf{x}^a \big\|^2 \Big\}}{(2\pi\epsilon/\beta)^{\frac{\beta m(p+1)}{2}}} \Big], \tag{B.13}$$

in which $(\mathbf{O}, \mathbf{U})$ is uniformly sampled over $\mathcal{U}_\beta(m) \times \mathcal{U}_\beta(n)$. Note that when $\mathbf{U}$ is uniformly distributed over $\mathcal{U}_\beta(n)$, the set of vectors $\{\mathbf{U}\mathbf{x}^a\}_{a=0}^p$ is uniformly distributed over the set of $(p+1)$ vectors in $\mathbb{K}^n$ with overlap matrix $\mathbf{Q}^x$. There is a completely similar result for $\mathbf{z}$ as well. The consequence is that we can replace in eq. (B.13) the average over $\mathbf{O}, \mathbf{U}$ by an average over

the vectors satisfying this constraint:

$$I_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) \tag{B.14}$$
$$\simeq \frac{1}{n}\ln \mathbb{E}_{\mathbf{\Phi}} \frac{\int_{\mathbb{K}}\prod_a \mathrm{d}\mathbf{x}^a\ \mathrm{d}\mathbf{z}^a\Big[\prod_{a\leq b}\delta(nQ^x_{ab}-(\mathbf{x}^a)^\dagger\mathbf{x}^b)\delta(mQ^z_{ab}-(\mathbf{z}^a)^\dagger\mathbf{z}^b)\Big]\frac{e^{-\frac{\beta}{2\epsilon}\sum_a\|\mathbf{z}^a-\frac{1}{\sqrt{n}}\mathbf{\Phi}\mathbf{x}^a\|^2}}{(2\pi\epsilon/\beta)^{\beta m(p+1)/2}}}{\int_{\mathbb{K}}\prod_a \mathrm{d}\mathbf{x}^a\ \mathrm{d}\mathbf{z}^a\Big[\prod_{a\leq b}\delta(nQ^x_{ab}-(\mathbf{x}^a)^\dagger\mathbf{x}^b)\delta(mQ^z_{ab}-(\mathbf{z}^a)^\dagger\mathbf{z}^b)\Big]}.$$

The numerator and the denominator correspond to two terms, that we denote $I_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) = I^{(n)}_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) - I^{(d)}_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z)$. We can introduce the Fourier-transform of the delta distribution to compute both terms, as in the previous sections. Let us start with the denominator. It reduces after Fourier-transformation to a Gaussian integral involving a block-diagonal matrix:

$$I^{(d)}_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) \simeq \frac{\beta}{2}\inf_{\mathbf{\Gamma}^x,\mathbf{\Gamma}^z}\Big[\mathrm{Tr}[\mathbf{Q}^x\mathbf{\Gamma}^x]+\alpha\mathrm{Tr}[\mathbf{Q}^z\mathbf{\Gamma}^z]-\ln\det\frac{\beta\mathbf{\Gamma}^x}{2\pi}-\alpha\ln\det\frac{\beta\mathbf{\Gamma}^z}{2\pi}\Big],$$

with symmetric (Hermitian) positive matrices $\mathbf{\Gamma}^x,\mathbf{\Gamma}^z$ of size $(p+1)$. The infimum is readily solved by $\mathbf{\Gamma}^x=(\mathbf{Q}^x)^{-1}$ and $\mathbf{\Gamma}^z=(\mathbf{Q}^z)^{-1}$, which yields:

$$I^{(d)}_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) \simeq \frac{\beta(\alpha+1)(p+1)}{2}(1+\ln\frac{2\pi}{\beta})+\frac{\beta}{2}\ln\det\mathbf{Q}^x+\frac{\alpha\beta}{2}\ln\det\mathbf{Q}^z. \tag{B.15}$$

Let us now compute the numerator with the same technique. We obtain:

$$I^{(n)}_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) \simeq \frac{\beta(p+1)}{2}\ln\frac{2\pi}{\beta\epsilon^\alpha}+\frac{\beta}{2}\inf_{\mathbf{\Gamma}^x,\mathbf{\Gamma}^z}\Big[\mathrm{Tr}[\mathbf{Q}^x\mathbf{\Gamma}^x]+\alpha\mathrm{Tr}[\mathbf{Q}^z\mathbf{\Gamma}^z]-\frac{1}{n}\ln\det\mathbf{M}_n\Big], \tag{B.16}$$

with a Hermitian block-matrix $\mathbf{M}_n$ that we write here in the tensor product form:

$$\mathbf{M}_n \equiv \begin{pmatrix}(\mathbf{\Gamma}^z+\frac{1}{\epsilon}\mathrm{I}_{p+1})\otimes\mathrm{I}_m & \frac{1}{\epsilon}\mathrm{I}_{p+1}\otimes\frac{\mathbf{\Phi}}{\sqrt{n}}\\ \frac{1}{\epsilon}\mathrm{I}_{p+1}\otimes\frac{\mathbf{\Phi}^\dagger}{\sqrt{n}} & \mathbf{\Gamma}^x\otimes\mathrm{I}_n+\frac{1}{\epsilon}\mathrm{I}_{p+1}\otimes\frac{\mathbf{\Phi}^\dagger\mathbf{\Phi}}{n}\end{pmatrix}.$$

Using the block-matrix determinant calculation

$$\det\begin{pmatrix}A & B\\ C & D\end{pmatrix} = \det A\times\det(D-CA^{-1}B),$$

we reach:

$$\begin{aligned}\frac{1}{n}\ln\det\mathbf{M}_n &= \alpha\ln\det\Big(\mathbf{\Gamma}^z+\frac{1}{\epsilon}\mathrm{I}_{p+1}\Big)\\ &\quad+\frac{1}{n}\ln\det\Big(\mathbf{\Gamma}^x\otimes\mathrm{I}_n+\frac{1}{\epsilon}\mathrm{I}_{p+1}\otimes\frac{\mathbf{\Phi}^\dagger\mathbf{\Phi}}{n}-\frac{1}{\epsilon^2}\Big(\mathbf{\Gamma}^z+\frac{1}{\epsilon}\mathrm{I}_{p+1}\Big)^{-1}\otimes\frac{\mathbf{\Phi}^\dagger\mathbf{\Phi}}{n}\Big),\\ &= (\alpha-1)\ln\det\Big(\mathbf{\Gamma}^z+\frac{1}{\epsilon}\mathrm{I}_{p+1}\Big)+\frac{1}{n}\ln\det\Big(\mathbf{\Gamma}^x\mathbf{\Gamma}^z\otimes\mathrm{I}_n+\frac{1}{\epsilon}\mathbf{\Gamma}^x\otimes\mathrm{I}_n+\frac{1}{\epsilon}\mathbf{\Gamma}^z\otimes\frac{\mathbf{\Phi}^\dagger\mathbf{\Phi}}{n}\Big),\\ &= (\alpha-1)\ln\det\Big(\mathbf{\Gamma}^z+\frac{1}{\epsilon}\mathrm{I}_{p+1}\Big)+\Big\langle\ln\det\Big(\mathbf{\Gamma}^x\mathbf{\Gamma}^z+\frac{1}{\epsilon}(\mathbf{\Gamma}^x+\lambda\mathbf{\Gamma}^z)\Big)\Big\rangle_\nu+\mathcal{O}_n(1),\end{aligned}$$

with λ distributed according to ν, the LSD of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$. This allows to write $I^{(n)}_{\rm int}$ from eq. (B.16) and to take the $\epsilon\downarrow 0$ limit, keeping the terms that do not vanish:

$$I^{(n)}_{\rm int}(p,\mathbf{Q}^x,\mathbf{Q}^z) \simeq \frac{\beta}{2}\inf_{\mathbf{\Gamma}^x,\mathbf{\Gamma}^z}[\mathrm{Tr}[\mathbf{Q}^x\mathbf{\Gamma}^x]+\alpha\mathrm{Tr}[\mathbf{Q}^z\mathbf{\Gamma}^z]-\langle\ln\det(\mathbf{\Gamma}^x+\lambda\mathbf{\Gamma}^z)\rangle_\nu]. \tag{B.17}$$

Finally, we again consider a replica-symmetric assumption for $\mathbf{\Gamma}^x, \mathbf{\Gamma}^z$, in the form:

$$\mathbf{\Gamma}^x = \begin{pmatrix} \Gamma_x & -\gamma_x & \cdots & -\gamma_x \\ -\gamma_x & \Gamma_x & \cdots & -\gamma_x \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma_x & -\gamma_x & \cdots & \Gamma_x \end{pmatrix}, \qquad \mathbf{\Gamma}^z = \begin{pmatrix} \Gamma_z & -\gamma_z & \cdots & -\gamma_z \\ -\gamma_z & \Gamma_z & \cdots & -\gamma_z \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma_z & -\gamma_z & \cdots & \Gamma_z \end{pmatrix}.$$

Again, by hermiticity, we have $\gamma_x, \gamma_z \in \mathbb{R}$. Combining eqs. (B.15) and (B.17) and using the replica symmetric assumption, we finally obtain the cumbersome expression:

$$\begin{aligned} \frac{2}{\beta} I_{\text{int}}(p, \mathbf{Q}_x, \mathbf{Q}_z) = \inf_{\Gamma_x, \gamma_x, \Gamma_z, \gamma_z} & [(p+1)Q_x\Gamma_x - p(p+1)q_x\gamma_x + \alpha(p+1)Q_z\Gamma_z - \alpha p(p+1)q_z\gamma_z \\ & - p\langle \ln(\Gamma_x + \gamma_x + \lambda\Gamma_z + \lambda\gamma_z)\rangle_\nu - \langle \ln[\Gamma_x - p\gamma_x + \lambda(\Gamma_z - p\gamma_z)]\rangle_\nu] - (\alpha+1)(p+1)\ln 2\pi e/\beta \\ & + (p+1)\ln\frac{2\pi}{\beta} - p\ln(Q_x - q_x) - \ln(Q_x + pq_x) - \alpha p \ln(Q_z - q_z) - \alpha \ln(Q_z + pq_z). \end{aligned} \tag{B.18}$$

A note on quenched and annealed averages – Importantly, we did not consider the average over $\mathbf{\Phi}$ to compute I_{int}. Indeed, the result only depends on the eigenvalue distribution of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$, which (by hypothesis) has large deviations in a scale at least $n^{1+\eta}$ with $\eta > 0$. Since we are looking at a scale exponential in n, we can thus consider that this eigenvalue distribution is equal to its limit value ν. However, one must be careful that this argument breaks down if our result starts to be sensitive to the extremal eigenvalues of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$. Since these variables typically have large deviations in the scale n (for instance for Wigner or Wishart matrices [DM06]), this could invalidate our calculation. This phenomenon is well-known in the study of "HCIZ" spherical integrals, and we gave an example of it in Section 1.5.3. We argue in Section B.2.4 that this possible issue, not discussed in [TK20], never arises for physical values of the overlaps.

Expressing the p-th moment

Combining the three previous results we finally obtain:

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\, \mathcal{Z}_n(\mathbf{Y})^p = \sup_{\substack{Q_x, q_x \\ Q_z, q_z}} [I_0(p, Q_x, q_x) + \alpha I_{\text{out}}(p, Q_z, q_z) + I_{\text{int}}(p, Q_x, q_x, Q_z, q_z)], \tag{B.19}$$

in which the three terms are given by eqs. (B.9),(B.10),(B.18).

B.2.3 The $p \downarrow 0$ limit

One can easily see that the RHS of eq. (B.19) is analytic in p. The next step of the replica method is to analytically extend this expression to arbitrary $p > 0$, before considering the limit $p \downarrow 0$.

Consistency of the limit

One must be careful that, when extending our expression to arbitrarily small $p > 0$, we satisfy the trivial condition $\lim_{p\downarrow 0} \ln \mathbb{E}\mathcal{Z}_n^p = 0$. As we will see, this will yield constraints on the diagonals

of the overlap matrices. Taking the limit $p \downarrow 0$ in the three terms of eq. (B.19) yields:

$$\begin{cases} I_0(0, Q_x, q_x) & = \inf_{\hat{Q}_x} \Big\{ \frac{\beta}{2} Q_x \hat{Q}_x + \ln \int_{\mathbb{K}} P_0(\mathrm{d}x) e^{-\frac{\beta \hat{Q}_x}{2}|x|^2} \Big\}, \\ I_{\text{out}}(0, Q_z, q_z) & = \inf_{\hat{Q}_z} \Big\{ \frac{\beta}{2} Q_z \hat{Q}_z + \frac{\beta}{2} \ln\Big(\frac{2\pi}{\beta \hat{Q}_z}\Big) \Big\}, \\ I_{\text{int}}(0, Q_x, q_x, Q_z, q_z) & = \inf_{\Gamma_x, \Gamma_z} \Big[\frac{\beta}{2} Q_x \Gamma_x + \frac{\alpha\beta}{2} Q_z \Gamma_z - \frac{\beta}{2} \langle \ln[\Gamma_x + \lambda \Gamma_z] \rangle_\nu \Big] \\ & \quad - \frac{\beta(\alpha+1)}{2}(1 + \ln \frac{2\pi}{\beta}) + \frac{\beta}{2} \ln \frac{2\pi}{\beta} - \frac{\beta}{2} \ln Q_x - \frac{\alpha\beta}{2} \ln Q_z. \end{cases}$$

One can easily solve the saddle point equations on $Q_z, \hat{Q}_z$, they give $\Gamma_z = 0$ and $\hat{Q}_z = 1/Q_z$. One can then find all the remaining variables easily: $Q_x = \rho$, $\hat{Q}_x = 0$, $\Gamma_x = \rho^{-1}$, $Q_z = \rho \langle \lambda \rangle_\nu / \alpha$, $\hat{Q}_z = 1/Q_z$, $\Gamma_z = 0$. This yields (we drop the vacuous dependency on q_x, q_z):

$$\begin{cases} I_0(0, Q_x = \rho) & = 0, \\ I_{\text{out}}\Big(0, Q_z = \dfrac{\rho \langle \lambda \rangle_\nu}{\alpha}\Big) & = \dfrac{\beta}{2} + \dfrac{\beta}{2} \ln\Big(\dfrac{2\pi \rho \langle \lambda \rangle_\nu}{\beta \alpha}\Big), \\ I_{\text{int}}\Big(0, Q_x = \rho, Q_z = \dfrac{\rho \langle \lambda \rangle_\nu}{\alpha}\Big) & = -\dfrac{\beta\alpha}{2}\Big(1 + \ln \dfrac{2\pi}{\beta}\Big) - \dfrac{\alpha\beta}{2} \ln \dfrac{\rho \langle \lambda \rangle_\nu}{\alpha}. \end{cases} \tag{B.20}$$

Recall that we have

$$\lim_{p \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E} \mathcal{Z}_n(\mathbf{Y})^p = I_0 + \alpha I_{\text{out}} + I_{\text{int}},$$

so that we obtain from eq. (B.20) that indeed the limit is consistent.

The replica-symmetric result

Using eq. (B.19) for the p-th moment and the consistency conditions we just derived, we obtain after using the replica trick:

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_n(\mathbf{Y}) = \sup_{q_x, q_z} [I_0(q_x) + \alpha I_{\text{out}}(q_z) + I_{\text{int}}(q_x, q_z)], \tag{B.21}$$

with the auxiliary functions:

$$\begin{aligned} I_0(q_x) &\equiv \inf_{\hat{q}_x \geq 0} \Big[-\frac{\beta \hat{q}_x q_x}{2} + \int_{\mathbb{K}} \mathcal{D}_\beta \xi P_0(\mathrm{d}x) e^{-\frac{\beta \hat{q}_x}{2}|x|^2 + \beta \sqrt{\hat{q}_x} x \cdot \xi} \ln \int_{\mathbb{K}} P_0(\mathrm{d}x) e^{-\frac{\beta \hat{q}_x}{2}|x|^2 + \beta \sqrt{\hat{q}_x} x \cdot \xi} \Big], \\ I_{\text{out}}(q_z) &\equiv \inf_{\hat{q}_z \geq 0} \Big\{ -\frac{\beta \hat{q}_z q_z}{2} - \frac{\beta}{2} \ln(\hat{Q}_z + \hat{q}_z) + \frac{\beta \hat{q}_z}{2\hat{Q}_z} + \int \mathrm{d}y \mathcal{D}_\beta \xi \; J(\hat{q}_z, y, \xi) \ln J(\hat{q}_z, y, \xi) \Big\}, \\ I_{\text{int}}(q_x, q_z) &\equiv \inf_{\gamma_x, \gamma_z \geq 0} \Big[\frac{\beta}{2}(\rho - q_x)\gamma_x + \frac{\alpha\beta}{2}(Q_z - q_z)\gamma_z - \frac{\beta}{2} \langle \ln(\rho^{-1} + \gamma_x + \lambda \gamma_z) \rangle_\nu \Big] \\ &\quad - \frac{\beta}{2} \ln(\rho - q_x) - \frac{\beta q_x}{2\rho} - \frac{\alpha\beta}{2} \ln(Q_z - q_z) - \frac{\alpha\beta q_z}{2Q_z}, \end{aligned}$$

with $Q_z \equiv \rho \langle \lambda \rangle_\nu / \alpha$ and $\hat{Q}_z = 1/Q_z$. Moreover, the domain of the supremum is $q_x \in [0, \rho]$ and $q_z \in [0, Q_z]$. The function $J(\hat{q}_z, y, \xi)$ appearing in the expression of I_{out} is defined as:

$$J(\hat{q}_z, y, \xi) \equiv \int_{\mathbb{K}} \mathcal{D}_\beta z P_{\text{out}}\Big(y \Big| \frac{z}{\sqrt{\hat{Q}_z + \hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z + \hat{q}_z)}} \xi \Big).$$

Note that compared to the calculation presented in the previous sections, we moved a term $(\beta\alpha/2)(1+\ln 2\pi/\beta)$ between I_{out} and I_{int}, and we also made a few straightforward change of variables in the expression of I_{out}. This is exactly the result given in Conjecture 6.1!

B.2.4 Concentration of the spectrum of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ and the absence of saturation

As emphasized during the computation of I_{int}, our calculation assumed that the extremization equations on (γ_x, γ_z) always admitted a solution. Moreover, we assumed that this solution is not sensitive to the extremal eigenvalues of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$. If this assumption is indeed true, the rate of the large deviations of the spectrum of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ was assumed to be large enough to justify our calculation. This important condition can be phrased by saying that for all physical values of (q_x, q_z), we must not "touch" the edge of the spectrum in the variational expression of I_{int}:

$$\frac{1}{\rho} + \gamma_x + \gamma_z \lambda_{\min}(\nu) > 0. \tag{B.22}$$

We justify here eq. (B.22) for all physical values of (q_x, q_z). We will combine three arguments:

(i) In the computation of I_{int} the matrix $\mathbf{\Gamma}^z$ is assumed to be Hermitian positive in the $p \downarrow 0$ limit. Since $\Gamma_z = 0$, this implies that we must have $\lambda_z \geq 0$.

(ii) The saddle point equation on q_x yields[1]:

$$\hat{q}_x = \frac{q_x}{\rho(\rho - q_x)} - \gamma_x. \tag{B.23}$$

(iii) Finally, we will derive a lower bound on q_x. As one can see from the computation of I_0, q_x is the optimal overlap achievable in the following scalar inference problem [BKM+19]:

$$Y_0 = \sqrt{\hat{q}_x} X^\star + Z, \tag{B.24}$$

in which one observes Y_0, and the noise Z is distributed according to $\mathcal{N}_\beta(0,1)$. The optimal estimator is given by the average of x under the posterior distribution, whose density is proportional to $P_0(x)e^{-\frac{\beta}{2}|y-\sqrt{\hat{q}_x}x|^2}$. If this is untractable for generic P_0, we can consider a suboptimal estimation by using a Gaussian prior with variance ρ in the estimation procedure (so that the problem is mismatched). This yields the bound:

$$q_x \geq \int \mathcal{D}_\beta \xi \frac{\left[\int_{\mathbb{K}} P_0(\mathrm{d}x)\; x\; e^{-\frac{\beta\hat{q}_x}{2}|x|^2+\beta\sqrt{\hat{q}_x}x\cdot\xi}\right]\cdot\left[\int_{\mathbb{K}} \mathrm{d}x\; x\; e^{-\frac{\beta|x|^2}{2\rho}}\; e^{-\frac{\beta\hat{q}_x}{2}|x|^2+\beta\sqrt{\hat{q}_x}x\cdot\xi}\right]}{\int_{\mathbb{K}} \mathrm{d}x\; e^{-\frac{\beta|x|^2}{2\rho}}\; e^{-\frac{\beta\hat{q}_x}{2}|x|^2+\beta\sqrt{\hat{q}_x}x\cdot\xi}}.$$

This can easily be simplified by performing the Gaussian integral, and yields the bound:

$$q_x \geq \frac{\rho^2\hat{q}_x}{1+\rho\hat{q}_x}. \tag{B.25}$$

Combining *(ii)* and *(iii)* gives:

$$q_x \geq \rho - \frac{\rho - q_x}{1 - \gamma_x(\rho - q_x)}.$$

[1] This relation is valid even if λ_x would "saturate" to a constant value that does not depend on (q_x, q_z).

Since $q_x \in [0, \rho]$, this implies in particular that $\gamma_x \geq 0$. Using this along with (i), we reach:

$$\frac{1}{\rho} + \gamma_x + \gamma_z \lambda_{\min}(\nu) \geq \frac{1}{\rho} > 0,$$

which is what we wanted to show.

Appendix C

Proving the replica formula: details in the committee machine

C.1 Positivity of some matrices

We first prove the following lemma which provides basic properties of the overlap matrix:

Lemma C.1 (*Positivity of some matrices*)

The matrices ρ, $\mathbb{E}\langle Q\rangle$ and $\rho - \mathbb{E}\langle Q\rangle$ are all positive semi-definite, i.e. in $\mathcal{S}_K^+$.

Proof of Lemma C.1 – The statement for ρ follows from its very definition. Note for further use that we have $\rho = n^{-1}\sum_{i=1}^n \mathbb{E}[W_i^*(W_i^*)^\intercal]$. Since by definition $Q = n^{-1}\sum_{i=1}^n W_i^* w_i^\intercal$, by the Nishimori identity (Proposition 1.1), we have

$$\mathbb{E}\langle Q\rangle = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\langle W_i^* w_i^\intercal\rangle = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\langle w_i\rangle\langle w_i^\intercal\rangle],$$

which is obviously in $\mathcal{S}_K^+$. Finally we note that

$$\rho - \mathbb{E}[\langle Q\rangle] = \frac{1}{n}\sum_{i=1}^n \Big(\mathbb{E}[W_i^*(W_i^*)^\intercal] - \mathbb{E}[\langle w_i\rangle\langle w_i^\intercal\rangle]\Big) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(W_i^* - \langle w_i\rangle)((W_i^*)^\intercal - \langle w_i^\intercal\rangle)]$$

where the last equality follows again from the Nishimori identity. This last expression is obviously in $\mathcal{S}_K^+$, i.e. $\mathbb{E}\langle Q\rangle \in \mathcal{S}_K^+(\rho)$. □

C.2 Properties of the auxiliary channels

Lemma C.2 (*Properties of ψ_{P_0}*)

Recall that ψ_{P_0} is the free entropy of the auxiliary channel of eq. (4.3). More precisely we have, for any $r \in \mathcal{S}_K^+$:

$$\psi_{P_0}(r) \equiv \mathbb{E}\ln\int_{\mathbb{R}^K} \mathrm{d}w\, P_0(w)\, e^{Y_0^\intercal r^{1/2} w - \frac{1}{2}w^\intercal r w}.$$

Then ψ_{P_0} is convex and differentiable on $\mathcal{S}_K^+$, and moreover $\nabla\psi_{P_0}(r) \in \mathcal{S}_K^+$ for any $r \in \mathcal{S}_K^+$.

Proof of Lemma C.2 – One can easily compute (either directly, or by the I-MMSE theorem for vector channels [RPD18]) $\nabla\psi_{P_0}(r) = (\rho - \mathbb{E}[\langle w\rangle\langle w\rangle^\intercal])/2$. Using the Nishimori Proposition 1.1, we can write it as $\nabla\psi_{P_0}(r) = \mathbb{E}[(w - \langle w\rangle)(w - \langle w\rangle)^\intercal]/2$, which is clearly a positive matrix. By a very similar computation (see for instance Lemma 4 of [RPD18]), one can check that $\nabla^2\psi_{P_0}$ is a positive operator on $\mathcal{S}_K^+ \times \mathcal{S}_K^+$, so that ψ_{P_0} is convex, which ends the proof. □

Lemma C.3 (*Properties of* Ψ_{out})

Recall that Ψ_{out} is the free entropy of the auxiliary channel of eq. (4.4). More precisely, for $q \in \mathcal{S}_K^+(\rho)$, we have:

$$\Psi_{\text{out}}(q) \equiv \mathbb{E} \ln \int_{\mathbb{R}^K} \mathrm{d}w \, \frac{e^{-\frac{1}{2}\|w\|^2}}{(2\pi)^{K/2}} \, P_{\text{out}}(\widetilde{Y}_0 | q^{1/2} V + (\rho - q)^{1/2} w).$$

Then Ψ_{out} is continuous and convex on $\mathcal{S}_K^+(\rho)$, and twice differentiable inside $\mathcal{S}_K^+(\rho)$. Moreover, for all $q \in \mathcal{S}_K^+(\rho)$, one has $\nabla \Psi_{\text{out}}(q) \in \mathcal{S}_K^+$.

Proof of Lemma C.3 – The continuity and differentiability of Ψ_{out} is easy, and the reader can simply follow the proof of Proposition 18 of [BKM+19]. It follows from H.2 which allows to use continuity and differentiation under the expectation, as all usual domination hypotheses are then easily verified.
One can compute the gradient and Hessian matrix of $\Psi_{\text{out}}(q)$, for q inside $\mathcal{S}_K^+(\rho)$, using Gaussian integration by parts and the Nishimori identity 1.1. The calculation is tedious and essentially follows the steps of Proposition 11 of [BKM+19]. Recall that $u_{\widetilde{Y}_0}(x) \equiv \ln P_{\text{out}}(\widetilde{Y}_0|x)$. We define the average $\langle - \rangle_{\text{sc}}$ (where sc stands for "scalar channel") as

$$\langle g(w) \rangle_{\text{sc}} \equiv \frac{\int_{\mathbb{R}^K} \mathcal{D}w P_{\text{out}}(\widetilde{Y}_0 | (\rho - q)^{1/2} w + q^{1/2} V) g(w)}{\int_{\mathbb{R}^K} \mathcal{D}w P_{\text{out}}(\widetilde{Y}_0 | (\rho - q)^{1/2} w + q^{1/2} V)}.$$

Using this definition, one arrives at:

$$\nabla \Psi_{\text{out}}(q) = \frac{1}{2} \mathbb{E} \Big\langle \nabla u_{\widetilde{Y}_0}((\rho - q)^{1/2} W^* + q^{1/2} V) \nabla u_{\widetilde{Y}_0}((\rho - q)^{1/2} w + q^{1/2} V)^\intercal \Big\rangle_{\text{sc}}.$$

Note that this gradient is actually a symmetric matrix of size K, as q is itself a matrix of size K. The Hessian $\nabla \nabla^\intercal \Psi_{\text{out}}$ with respect to q is thus a 4-tensor that can be computed similarly:

$$\begin{aligned} \nabla \nabla^\intercal \Psi_{\text{out}}(q) = \frac{1}{2} \mathbb{E} \Big[\Big(\Big\langle \frac{\nabla \nabla^\intercal P_{\text{out}}(\widetilde{Y}_0 | (\rho - q)^{1/2} w + q^{1/2} V)}{P_{\text{out}}(\widetilde{Y}_0 | (\rho - q)^{1/2} w + q^{1/2} V)} \Big\rangle_{\text{sc}} \\ - \Big\langle \nabla u_{\widetilde{Y}_0}((\rho - q)^{1/2} W^* + q^{1/2} V) \nabla u_{\widetilde{Y}_0}((\rho - q)^{1/2} w + q^{1/2} V)^\intercal \Big\rangle_{\text{sc}} \Big)^{\otimes 2} \Big]. \end{aligned}$$

In this expression, $\otimes 2$ means the "tensorized square" of a matrix, i.e. for any matrix M of size $K \times K$, $M^{\otimes 2}$ is a 4-tensor with indices $M^{\otimes 2}_{l_0 l_1 l_2 l_3} = M_{l_0 l_1} M_{l_2 l_3}$. From this expression, it is clear that the Hessian of Ψ_{out} is always positive, when seen as a matrix with rows and columns in $\mathcal{S}_K$, and thus Ψ_{out} is convex, which ends the proof. □

C.3 Setting in the Hamiltonian language

In this section, we set up some notations which will be useful in the following Section C.4. Let $u_y(x) \equiv \ln P_{\text{out}}(y|x)$. Here $x \in \mathbb{R}^K$ and $y \in \mathbb{R}$. We will denote by $\nabla u_y(x)$ the K-dimensional gradient w.r.t. x, and $\nabla \nabla^\intercal u_y(x)$ the $K \times K$ Hessian w.r.t. x. Moreover $\nabla P_{\text{out}}(y|x)$ and $\nabla \nabla^\intercal P_{\text{out}}(y|x)$ also denote the K-dimensional gradient and Hessian w.r.t. x. We will also use the matrix identity

$$\nabla \nabla^\intercal u_{Y_\mu}(x) + \nabla u_{Y_\mu}(x) \nabla^\intercal u_{Y_\mu}(x) = \frac{\nabla \nabla^\intercal P_{\text{out}}(Y_\mu|x)}{P_{\text{out}}(Y_\mu|x)}. \tag{C.1}$$

Finally we will use the matrices $\mathbf{w} \in \mathbb{R}^{n\times K}$, $\mathbf{u} \in \mathbb{R}^{m\times K}$, $\mathbf{Y}_t \in \mathbb{R}^m$, $\mathbf{Y}'_t \in \mathbb{R}^{n\times K}$, $\mathbf{X} \in \mathbb{R}^{m\times n}$, $\mathbf{V} \in \mathbb{R}^{m\times K}$, $\mathbf{W}^* \in \mathbb{R}^{n\times K}$ and $\mathbf{U}^* \in \mathbb{R}^{m\times K}$. It is convenient to reformulate the expression of the free entropy $f_{n,\epsilon}(t)$ in the Hamiltonian language. We introduce an *interpolating Hamiltonian*:

$$\mathcal{H}_t(\mathbf{w},\mathbf{u};\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V}) \equiv -\sum_{\mu=1}^{m} u_{Y_{t,\mu}}(s_{t,\mu}) + \frac{1}{2}\sum_{i=1}^{n} \|Y'_{t,i} - R_1(t)^{1/2}\, w_i\|_2^2, \tag{C.2}$$

where we recall that $s_{t,\mu}$ is defined in eq. (4.13). The expression of $\mathcal{H}_t(\mathbf{W}^*,\mathbf{U}^*;\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})$ is similar to eq. (C.2), but with $\mathbf{w}$ replaced by $\mathbf{W}^*$ and $s_{t,\mu}$ replaced by $S_{t,\mu}$ given by eq. (4.11). The average free entropy at time t (cf eq. (4.15)) then reads

$$f_{n,\epsilon}(t) \equiv \frac{1}{n}\mathbb{E}\ln \int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{w}\, P_0(\mathbf{w}) \int_{\mathbb{R}^{m\times K}} \mathcal{D}\mathbf{u}\, e^{-\mathcal{H}_t(\mathbf{w},\mathbf{u};\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})}.$$

Note that to develop the calculations it is sometimes fruitful to represent the expectations over $\mathbf{W}^*, \mathbf{U}$ explicitly as:

$$f_{n,\epsilon}(t) = \frac{1}{n}\mathbb{E}_{\mathbf{X},\mathbf{V},\mathbf{Y}_t,\mathbf{Y}'_y} \int \mathrm{d}P_0(\mathbf{W}^*)\mathcal{D}\mathbf{U}^* e^{-\mathcal{H}_t(\mathbf{W}^*,\mathbf{U};\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})} \ln \int \mathrm{d}P_0(\mathbf{w})\mathcal{D}\mathbf{u}\, e^{-\mathcal{H}_t(\mathbf{w},\mathbf{u};\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})}.$$

C.4 Free entropy variation: Proof of Proposition 4.3

The proof provided here follows very closely the one in [BKM+19] for the case $K = 1$, so that we will sometimes refer to this paper for more details. We first prove that for all $t \in (0,1)$:

$$\begin{aligned}\frac{\mathrm{d}f_{n,\epsilon}(t)}{\mathrm{d}t} = &-\frac{1}{2}\mathbb{E}\Big\langle \mathrm{Tr}\Big[\Big(\frac{1}{n}\sum_{\mu=1}^{m} \nabla u_{Y_{t,\mu}}(s_{t,\mu})\nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal - r(t)\Big)\Big(\frac{1}{n}\sum_{i=1}^{n} W_i^* w_i^\intercal - q(t)\Big)\Big]\Big\rangle_{n,t,\epsilon} \\ &+ \frac{1}{2}\mathrm{Tr}[r(t)(q(t)-\rho)] - \frac{A_n}{2}\,,\end{aligned} \tag{C.3}$$

where

$$A_n \equiv \mathbb{E}\Big[\mathrm{Tr}\Big[\frac{1}{\sqrt{n}}\sum_{\mu=1}^{m} \frac{\nabla\nabla^\intercal P_{\mathrm{out}}(Y_{t,\mu}|S_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|S_{t,\mu})}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i^*(W_i^*)^\intercal - \rho)\Big)\Big]\frac{1}{n}\ln \mathcal{Z}_{n,\epsilon}(t)\Big]\,.$$

Once we will have proven eq. (C.3), we show that A_n goes to 0 as $n \to \infty$ uniformly in $t \in [0,1]$ in order to conclude the proof of Proposition 4.3. Recall that the Hamiltonian was defined in eq. (C.2). Its t-derivative evaluated at the ground-truth matrices is given by

$$\begin{aligned}&\frac{\mathrm{d}\mathcal{H}_t}{\mathrm{d}t}(\mathbf{W}^*,\mathbf{U}^*;\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V}) \\ &= -\sum_{\mu=1}^{m} \nabla^\intercal u_{Y_{t,\mu}}(S_{t,\mu})\frac{\mathrm{d}S_{t,\mu}}{\mathrm{d}t} - \sum_{i=1}^{n}\Big(\frac{\mathrm{d}R_1(t)^{1/2}}{\mathrm{d}t}W_i^*\Big)^\intercal (Y'_{t,i} - R_1(t)^{1/2}W_i^*), \\ &= -\sum_{\mu=1}^{m} \mathrm{Tr}\Big[\frac{\mathrm{d}S_{t,\mu}}{\mathrm{d}t}\nabla^\intercal u_{Y_{t,\mu}}(S_{t,\mu})\Big] - \sum_{i=1}^{n}\mathrm{Tr}\Big[\Big(\frac{\mathrm{d}R_1(t)^{1/2}}{\mathrm{d}t}\Big)^\intercal (Y'_{t,i} - R_1(t)^{1/2}W_i^*)W_i^{*\intercal}\Big].\end{aligned} \tag{C.4}$$

The t-derivative of $f_{n,\epsilon}(t)$ thus reads, for $t \in (0,1)$:

$$\frac{\mathrm{d}f_{n,\epsilon}(t)}{\mathrm{d}t} = \\ \underbrace{-\frac{1}{n}\mathbb{E}\Big[\frac{\mathrm{d}\mathcal{H}_t}{\mathrm{d}t}(\mathbf{W}^*,\mathbf{U}^*;\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})\ln\mathcal{Z}_{n,\epsilon}(t)\Big]}_{T_1} \underbrace{-\frac{1}{n}\mathbb{E}\Big\langle\frac{\mathrm{d}\mathcal{H}_t}{\mathrm{d}t}(\mathbf{w},\mathbf{u};\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})\Big\rangle_{n,t,\epsilon}}_{T_2}. \tag{C.5}$$

First, we note that $T_2 = 0$ by the Nishimori identity, Proposition 1.1:

$$T_2 = \frac{1}{n}\mathbb{E}\Big\langle\frac{\mathrm{d}\mathcal{H}_t}{\mathrm{d}t}(\mathbf{w},\mathbf{u};\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V})\Big\rangle_{n,t,\epsilon} = \frac{1}{n}\mathbb{E}\,\frac{\mathrm{d}\mathcal{H}_t}{\mathrm{d}t}(\mathbf{W}^*,\mathbf{U}^*;\mathbf{Y}_t,\mathbf{Y}'_t,\mathbf{X},\mathbf{V}) = 0\,.$$

We now compute T_1. We start from eq. (C.4) and we consider for the moment the first term (recall also the expression (4.11) for $S_{t,\mu}$):

$$\mathbb{E}\Big\{\mathrm{Tr}\Big[\frac{\mathrm{d}S_{t,\mu}}{\mathrm{d}t}\nabla^\intercal u_{Y_{t,\mu}}(S_{t,\mu})\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big\} = \mathbb{E}\Big[\mathrm{Tr}\Big[\Big\{-\frac{\sum_{i=1}^n X_{\mu i}W_i^*}{2\sqrt{n(1-t)}} \\ + \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{R_2(t)}V_\mu + \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{t\rho - R_2(t) + 2s_n\mathrm{I}_K}\,U_\mu^*\Big\}\nabla^\intercal u_{Y_{t,\mu}}(S_{t,\mu})\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big]. \tag{C.6}$$

We then compute the first term of the right-hand side of eq. (C.6). By Gaussian integration by parts w.r.t. $X_{\mu i}$ (recall hypothesis H.3), and using eq. (C.1), we find after some algebra

$$-\frac{1}{2\sqrt{n(1-t)}}\mathbb{E}\Big[\mathrm{Tr}\Big[\sum_{i=1}^n X_{\mu i}W_i^*\nabla^\intercal u_{Y_{t,\mu}}(S_{t,\mu})\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big] \\ = -\frac{1}{2}\mathbb{E}\Big[\mathrm{Tr}\Big[\frac{1}{n}\sum_{i=1}^n W_i^*W_i^\intercal\frac{\nabla\nabla^\intercal P_{\mathrm{out}}(Y_{t,\mu}|S_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|S_{t,\mu})}\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big] \\ -\frac{1}{2}\mathbb{E}\Big\langle\mathrm{Tr}\Big[\frac{1}{n}\sum_{i=1}^n W_i^*w_i^\intercal\nabla u_{Y_{t,\mu}}(S_{t,\mu})\nabla^\intercal u_{Y_{t,\mu}}(s_{t,\mu})\Big]\Big\rangle_{n,t,\epsilon}. \tag{C.7}$$

For the remaining terms of the right hand side of (C.6), we use again Gaussian integrations by parts but this time w.r.t. V_μ, U_μ^* which have i.i.d. $\mathcal{N}(0,1)$ entries. This calculation has to be done carefully using the cyclicity and linearity of the trace, and with the help of the identity

$$\frac{\mathrm{d}}{\mathrm{d}t}M(t) = \sqrt{M(t)}\,\frac{\mathrm{d}\sqrt{M(t)}}{\mathrm{d}t} + \frac{\mathrm{d}\sqrt{M(t)}}{\mathrm{d}t}\sqrt{M(t)} \tag{C.8}$$

for any $M(t) \in \mathcal{S}_K^+$. Applying eq. (C.8) to $\int_0^t q(s)ds$ and $\int_0^t(\rho - q(s))ds$, as well as the identity of eq. (C.1), we reach after some algebra

$$\mathbb{E}\Big[\mathrm{Tr}\Big[\Big(\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{R_2(t)}V_\mu + \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{t\rho - R_2(t) + 2s_n\mathrm{I}_K}\,U_\mu^*\Big)\nabla^\intercal u_{Y_\mu}(S_{\mu,t})\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big] \\ = \mathbb{E}\Big[\mathrm{Tr}\Big[\rho\frac{\nabla\nabla^\intercal P_{\mathrm{out}}(Y_{t,\mu}|S_{\mu,t})}{P_{\mathrm{out}}(Y_{t,\mu}|S_{\mu,t})}\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big] + \mathbb{E}\Big\langle\mathrm{Tr}\Big[q(t)\nabla u_{Y_{t,\mu}}(S_{\mu,t})\nabla^\intercal u_{Y_{t,\mu}}(s_{\mu,t})\Big]\Big\rangle_{n,t,\epsilon}. \tag{C.9}$$

This completes the computation of the terms of eq. (C.6). It now remains to compute the second term of the right hand side of eq. (C.4). Recall that $Y'_{t,i} - \sqrt{R_1(t)}W_i^* = Z'_i \sim \mathcal{N}(0,\mathrm{I}_K)$. Using Gaussian integration by parts as well as eq. (C.8) one obtains

$$\mathbb{E}\Big[\mathrm{Tr}\Big[\Big(\frac{\mathrm{d}\sqrt{R_1(t)}}{\mathrm{d}t}\Big)^\intercal(Y'_{t,i} - \sqrt{R_1(t)}W_i^*)W_i^{*\intercal}\Big]\ln\mathcal{Z}_{n,\epsilon}(t)\Big] = \mathrm{Tr}\Big[\sqrt{R_1(t)}(\mathbb{E}\langle W_i^*w_i^\intercal\rangle_{n,t,\epsilon} - \rho)\Big]. \tag{C.10}$$

Finally the term T_1 is obtained by putting together eqs. (C.6), (C.7), (C.9) and (C.10).

This ends the derivation of eq. (C.3). It now remains to check that $A_n \to 0$ as $n \to +\infty$ uniformly in $t \in [0,1]$. The proof from [BKM+19] can easily be adapted so we give here just a few indications for the ease of the reader. First one notices that

$$\mathbb{E}\Big[\frac{\nabla\nabla^\intercal P_{\rm out}(Y_{t,\mu}|S_{t,\mu})}{P_{\rm out}(Y_\mu|S_{t,\mu})}\,\Big|\,\mathbf{W}^*, \{S_{t,\mu}\}_{\mu=1}^m\Big] = \int \mathrm{d}Y_\mu \nabla\nabla^\intercal P_{\rm out}(Y_{t,\mu}|S_{t,\mu}) = 0,$$

so that by the tower property of the conditional expectation one gets

$$\mathbb{E}\Big\{\mathrm{Tr}\Big[\frac{1}{\sqrt{n}}\sum_{\mu=1}^m \frac{\nabla\nabla^\intercal P_{\rm out}(Y_{t,\mu}|S_{t,\mu})}{P_{\rm out}(Y_{t,\mu}|S_{t,\mu})}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_i^*(W_i^*)^\intercal - \rho)\Big)\Big]\Big\} = 0. \tag{C.11}$$

Next, one shows by standard second moment methods that $\mathbb{E}[(\ln \mathcal{Z}_{n,\epsilon}(t)/n - f_{n,\epsilon}(t))^2] \to 0$ as $n \to +\infty$ uniformly in $t \in [0,1]$ (see [BKM+19] for the proof at $K = 1$, that generalizes straightforwardly for any finite K). Then, using this last fact together with (C.11), and under hypotheses H.1, H.2, H.3, an easy application of the Cauchy-Schwarz inequality implies $A_n \to 0$ as $n \to +\infty$ uniformly in $t \in [0,1]$. This ends the proof of Proposition 4.3.

C.5 A few technical lemmas

Lemma C.4 (*Cauchy-Lipschitz Theorem and Liouville Formula*)

Let

$$F : \left|\begin{array}{ccc} [0,1]\times(0,+\infty)^d & \to & [0,+\infty)^d \\ (t,z) & \mapsto & F(t,z) \end{array}\right.$$

be a continuous bounded function. Assume that F admits continuous partial derivatives $\partial F/\partial z_i$ $(i = 1,\dots,d)$ on its domain of definition. Then, for all $\epsilon \in (0,+\infty)^d$, the Cauchy problem

$$y(0) = \epsilon \qquad \text{and} \qquad y'(t) = F\big(t, y(t)\big) \tag{C.12}$$

admits a unique solution $t \mapsto y(t,\epsilon)$. For all $t \in [0,1]$, the mapping $z_t : \epsilon \mapsto y(t,\epsilon)$ is a diffeomorphism of class $\mathcal{C}^1$, from $(0,+\infty)^d$ to $z_t((0,+\infty)^d)$. Moreover the determinant $J(z_t)(\epsilon)$ of the Jacobian of z_t at ϵ verifies

$$J(z_t)(\epsilon) = \det\Big(\Big(\frac{\partial y_i}{\partial \epsilon_j}\Big)_{i,j}\Big) = \exp\Big(\int_0^t \sum_{i=1}^d \frac{\partial F_i}{\partial z_i}(s, y(s,\epsilon))ds\Big). \tag{C.13}$$

Thus, in particular, if in addition $\sum_{i=1}^d (\partial F_i/\partial z_i) \geq 0$ then $J(z_t)(\epsilon) \geq 1$ for all ϵ.

Proof of Lemma C.4 – The existence and uniqueness of the solution of (C.12) follows from the classical Cauchy-Lipschitz Theorem. The solution is indeed defined on all the segment $[0,1]$ because F is bounded. Theorem 3.1 from Chapter 5 in [Har82] gives that y admits continuous partial derivatives $\partial_{\epsilon_i} y$ for $i = 1,\dots,d$, and Corollary 3.1 from Chapter 5 in the same reference states the Liouville formula (C.13). It remains to show that z_t is a $\mathcal{C}^1$ diffeomorphism. By the Cauchy-Lipschitz theorem, two solutions of $y'(t) = F(t, y(t))$ that are equal at some $t \in [0,1]$ are equal everywhere. This implies that the mapping $z_t : \epsilon \mapsto y(t,\epsilon)$ is injective, for all $t \in [0,1]$. Since y admits continuous partial derivatives in ϵ_i, $i = 1,\dots,d$, we obtain that z_t is of class $\mathcal{C}^1$ on $(0,+\infty)^d$. Now, the equation (C.13) gives that $J(z_t)(\epsilon) > 0$ for all $\epsilon \in (0,+\infty)^d$. The local inversion theorem gives then that z_t is a $\mathcal{C}^1$ diffeomorphism. □

Lemma C.5 (*Boundedness of an overlap fluctuation*)

Recall that for the sake of the proof we added (cf. Remark 4.1) a small Gaussian noise of variance $\Delta > 0$ to the observations. Under hypothesis H.2 one can then find a constant $C(\varphi, K, \Delta) < +\infty$ (independent of n, t, ϵ) such that for any $R_n \in \mathcal{S}_K^+$ we have

$$\mathbb{E}\Big\langle \Big\| \frac{1}{n}\sum_{\mu=1}^{m} \nabla u_{Y_{t,\mu}}(s_{t,\mu}) \nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal - R_n \Big\|_F^2 \Big\rangle_{n,t,\epsilon} \leq 2\mathrm{Tr}(R_n^2) + \alpha^2 C(\varphi, K, \Delta). \tag{C.14}$$

We note that the constant remains bounded as $\Delta \to 0$, and diverges as $K \to +\infty$.

Proof of Lemma C.5 – It is easy to see that for symmetric matrices A, B we have $\mathrm{Tr}(A-B)^2 \leq 2(\mathrm{Tr}A^2 + \mathrm{Tr}B^2)$. Therefore

$$\begin{aligned}\mathbb{E}\Big\langle \Big\| \frac{1}{n}\sum_{\mu=1}^{m} \nabla u_{Y_{t,\mu}}(s_{t,\mu}) \nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal - R_n \Big\|_F^2 \Big\rangle_{n,t,\epsilon} \\ \leq 2\mathrm{Tr}(R_n^2) + 2\mathbb{E}\Big\langle \mathrm{Tr}\Big(\frac{1}{n}\sum_{\mu=1}^{m} \nabla u_{Y_{t,\mu}}(s_{t,\mu}) \nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal \Big)^2 \Big\rangle_{n,t,\epsilon}.\end{aligned}$$

In the rest of the argument we bound the second term of the right hand side of this last inequality. Using the triangle inequality and then Cauchy-Schwarz we obtain

$$\begin{aligned}\mathbb{E}\Big\langle \Big\| \frac{1}{n}\sum_{\mu=1}^{m} \nabla u_{Y_{t,\mu}}(s_{t,\mu}) \nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal \Big\|_F^2 \Big\rangle_{n,t,\epsilon} &\leq \mathbb{E}\Big\langle \frac{1}{n^2}\Big(\sum_{\mu=1}^{m} \|\nabla u_{Y_{t,\mu}}(s_{t,\mu}) \nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal\|_F \Big)^2 \Big\rangle_{n,t,\epsilon} \\ &\leq \mathbb{E}\Big\langle \frac{1}{n^2}\Big(\sum_{\mu=1}^{m} \|\nabla u_{Y_{t,\mu}}(s_{t,\mu})\|_2 \|\nabla u_{Y_{t,\mu}}(S_{t,\mu})^\intercal\|_2 \Big)^2 \Big\rangle_{n,t,\epsilon}. \end{aligned} \tag{C.15}$$

Recall the random representation of the transition kernel:

$$u_{Y_{t,\mu}}(s) = \ln P_{\mathrm{out}}(Y_{t,\mu}|x) = \ln \int \mathrm{d}P_A(a_\mu) \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{1}{2\Delta}(Y_{t,\mu} - \varphi(x,a_\mu))^2},$$

and thus

$$\nabla u_{Y_{t,\mu}}(x) = \frac{\int \mathrm{d}P_A(a_\mu)(Y_{t,\mu} - \varphi(x,a_\mu)) \nabla\varphi(x,a_\mu) e^{-\frac{1}{2\Delta}(Y_{t,\mu} - \varphi(x,a_\mu))^2}}{\int \mathrm{d}P_A(a_\mu) e^{-\frac{1}{2\Delta}(Y_{t,\mu} - \varphi(x,a_\mu))^2}},$$

where $\nabla\varphi$ is the K-dimensional gradient w.r.t. the first argument $x \in \mathbb{R}^K$. From the observation model we get $|Y_{t,\mu}| \leq \sup|\varphi| + \sqrt{\Delta}|Z_\mu|$, where the supremum is taken over both arguments of φ, and thus we immediately obtain for all $s \in \mathbb{R}^K$

$$\|\nabla u_{Y_{t,\mu}}(x)\| \leq (2\sup|\varphi| + \sqrt{\Delta}|Z_\mu|) \sup\|\nabla\varphi\|. \tag{C.16}$$

From eqs. (C.15) and (C.16) we see that it suffices to check that

$$\frac{m^2}{n^2}\mathbb{E}\big[((2\sup|\varphi| + |Z_\mu|)^2 (\sup\|\nabla\varphi\|)^2)^2\big] \leq C(\varphi, K, \Delta),$$

where $C(\varphi, K, \Delta) < +\infty$ is a finite constant depending only on φ, K, and Δ. This is easily seen by expanding all squares and using that $m/n \to \alpha$. This ends the proof of Lemma C.5. □

Appendix D

Technical results of Part II

D.1 Generalization error in the committee machine

D.1.1 Bayes-optimal and Gibbs generalization error

We detail here two different possible definitions of the generalization error, and how they are related. Recall that we wish to estimate $\mathbf{W}^*$ from the observation of $\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)$. In the following, we denote $\mathbb{E}$ for the average over the (quenched) teacher weights $\mathbf{W}^*$ and the data $\mathbf{X}$, and $\langle - \rangle$ for the Gibbs average over the posterior distribution of $\mathbf{W}$. we can define two notions of generalization error, namely the *Gibbs generalization error* and the *Bayes-optimal generalization error* (which is the one considered in eq. (4.7)) as:

$$\begin{cases} \epsilon_g^{\text{Gibbs}} &\equiv \dfrac{1}{2}\mathbb{E}\big\langle[\varphi_{\text{out}}(\mathbf{X}\mathbf{W}) - \varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)]^2\big\rangle, \\ \epsilon_g^{\text{Bayes}} &\equiv \dfrac{1}{2}\mathbb{E}\big[\big(\langle\varphi_{\text{out}}(\mathbf{X}\mathbf{W})\rangle - \varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)\big)^2\big]. \end{cases} \tag{D.1}$$

Using the Nishimori identity 1.1, one can show that:

$$\begin{aligned} \epsilon_g^{\text{Bayes}} &= \frac{1}{2}\mathbb{E}[\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)^2] + \frac{1}{2}\mathbb{E}[\langle\varphi_{\text{out}}(\mathbf{X}\mathbf{W})\rangle^2] - \mathbb{E}\langle\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)\varphi_{\text{out}}(\mathbf{X}\mathbf{W})\rangle, \\ &= \frac{1}{2}\mathbb{E}[\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)^2] - \frac{1}{2}\mathbb{E}\langle\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)\varphi_{\text{out}}(\mathbf{X}\mathbf{W})\rangle. \end{aligned} \tag{D.2}$$

Using again the Nishimori identity one can write:

$$\epsilon_g^{\text{Gibbs}} = \mathbb{E}[\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)^2] - \mathbb{E}\langle\varphi_{\text{out}}(\mathbf{X}\mathbf{W}^*)\varphi_{\text{out}}(\mathbf{X}\mathbf{W})\rangle,$$

which shows that $\epsilon_g^{\text{Gibbs}} = 2\epsilon_g^{\text{Bayes}}$, and relates the two definitions of the generalization error.

D.1.2 The generalization error at $K = 2$ with committee symmetry

In this subsection alone, we consider the $K = 2$ case. From the definition of the generalization error that we saw in Section D.1.1, one can obtain an explicit expression in the $K = 2$ case. Recall that we assume *committee symmetry*, which here reads

$$q = \begin{pmatrix} q_d + \frac{q_a}{2} & \frac{q_a}{2} \\ \frac{q_a}{2} & q_d + \frac{q_a}{2} \end{pmatrix}.$$

For concision, we denote here $\text{sign}(x) = \sigma(x)$. One obtains from eq. (D.2):

$$\frac{1}{2} - 2\epsilon_g^{\text{Bayes,K=2}} = \int_{\mathbb{R}^4} \mathcal{D}x\, \sigma[\sigma(x_1) + \sigma(x_2)] \tag{D.3}$$
$$\times\, \sigma\Big\{\sigma\Big[(\frac{q_a}{2} + q_d)x_1 + \frac{q_a}{2}x_2 + x_3\sqrt{1 - \frac{q_a^2}{2} - q_a q_d - q_d^2}\Big]$$
$$+\, \sigma\Big[\frac{q_a}{2}x_1 + (\frac{q_a}{2} + q_d)x_2 - x_3\frac{q_a(q_d + \frac{q_a}{2})}{\sqrt{1 - \frac{q_a^2}{2} - q_a q_d - q_d^2}} + x_4\sqrt{\frac{(1 - q_d^2)(1 - (q_a + q_d)^2)}{1 - \frac{q_a^2}{2} - q_a q_d - q_d^2}}\Big]\Big\}.$$

These integrals were then computed using Monte-Carlo methods to obtain the generalization error in Fig. 4.2.

D.2 Large K limit in the committee machine

We consider the large K limit[1] for a sign activation function, and for different priors on the weights. We also consider a noiseless channel. We assume a committee symmetric solution, i.e. the matrices q and $\hat{q}$ (q and r in the notations of Theorem 4.1) are of the type $q = q_d \mathrm{I}_K + q_a \mathbf{1}_K \mathbf{1}_K^\intercal / K$, with the unit vector $\mathbf{1}_K \equiv (1)_{l=1}^K$, and similarly for $\hat{q}$.

In the large K limit, this scaling of the order parameters is natural. Indeed, assume that the covariance of the prior is $\rho = \mathrm{I}_K$. Since both q and $(\rho - q)$ are assumed to be positive matrices, it is easily shown to imply that $0 \leq q_d \leq 1$ and $0 \leq q_a + q_d \leq 1$, so that these variables stay of order $\mathcal{O}(1)$.

In the following of this section we take notations arising from the replica calculation of Section B.1. In particular, we note $Q^0 = \rho$, and we will take eq. (B.6) as the expression of the free entropy (recall that it is equivalent to Theorem 4.1, so that this expression is mathematically rigorous).

D.2.1 Limit of the channel integral

In the following, we consider $Q^0 = \sigma^2 \mathrm{I}_K$. We are interested here in computing the leading order term in the "channel" integral I_C of eq. (B.6). Note that replacing σ^2 by 1 in this equation only amounts to replacing q by q/σ^2, so we can assume $\sigma^2 = 1$ without loss of generality. We write I_C as $I_C = \sum_{y=\pm 1} \int_{\mathbb{R}^K} \mathcal{D}\xi\, I_C(y, \xi) \ln I_C(y, \xi)$, with the definition

$$I_C(y, \xi) \equiv \int_{\mathbb{R}^K} \mathcal{D}Z P_{\text{out}}\{y | (Q^0 - q)^{1/2} Z + q^{1/2}\xi\}.$$

Recall that we consider a sign activation function and no noise. Moreover, we assume committee symmetry (see the remark above). Note that this implies $q^{1/2} = \sqrt{q_d}\mathrm{I}_K + (\sqrt{q_a + q_d} - \sqrt{q_d})\mathbf{1}_K\mathbf{1}_K^\intercal/K$ and $(Q_0 - q)^{1/2} = \sqrt{1 - q_d}\mathrm{I}_K + (\sqrt{1 - q_a - q_d} - \sqrt{1 - q_d})\mathbf{1}_K\mathbf{1}_K^\intercal/K$. All together, this yields the following expression for $I_C(y, \xi)$:

$$I_C(y, \xi)$$
$$= \int_{\mathbb{R}^K} \mathcal{D}Z\, \delta\Big\{y - \text{sign}\Big[\frac{1}{\sqrt{K}}\sum_{l=1}^K \text{sign}\Big[\sqrt{1 - q_d}Z_l + (\sqrt{1 - q_a - q_d} - \sqrt{1 - q_d})\frac{\mathbf{1}_K^\intercal Z}{K} + (q^{1/2}\xi)_l\Big]\Big]\Big\}.$$

[1] Note that a similar limit has been derived in the context of coding with sparse superposition codes [BK17].

We introduce a new variable $w \equiv \mathbf{1}_K^\intercal Z/\sqrt{K}$ as well as another variable u being the argument of the outer sign function in the previous equations. Using the Fourier transform of the Dirac distribution one obtains:

$$I_C(y,\xi) = \int_{\mathbb{R}} \frac{\mathrm{d}w\mathrm{d}\hat{w}}{2\pi}\frac{\mathrm{d}u\mathrm{d}\hat{u}}{2\pi} e^{iw\hat{w}+iu\hat{u}}\delta_{y,\mathrm{sign}(u)} \\ \times \prod_{l=1}^{K}\int_{\mathbb{R}} \mathcal{D}z e^{-i\hat{w}\frac{z}{\sqrt{K}}} e^{-\frac{i\hat{u}}{\sqrt{K}}\mathrm{sign}\left[z+\left(\sqrt{\frac{1-q_a-q_d}{1-q_d}}-1\right)\frac{w}{\sqrt{K}}+\frac{1}{\sqrt{1-q_d}}(q^{1/2}\xi)_l\right]}.$$

We denote

$$\lambda_l(w,\xi) \equiv \left[\sqrt{\frac{1-q_a-q_d}{1-q_d}}-1\right]\frac{w}{\sqrt{K}} + \frac{1}{\sqrt{1-q_d}}(q^{1/2}\xi)_l.$$

For $1 \le l \le K$, one can rewrite the factorized integral in $I_C(y,\xi)$ as:

$$I_C(y,\xi) = \int_{\mathbb{R}} \frac{\mathrm{d}w\mathrm{d}\hat{w}}{2\pi}\frac{\mathrm{d}u\mathrm{d}\hat{u}}{2\pi} e^{iw\hat{w}+iu\hat{u}}\delta_{y,\mathrm{sign}(u)} \prod_{l=1}^{K} J(\lambda_l(w,\xi),\hat{w},\hat{u}), \tag{D.4}$$

in which we defined

$$J(\lambda_l(w,\xi),\hat{w},\hat{u}) \equiv e^{-\frac{\lambda_l^2}{2}+i\lambda_l\frac{\hat{w}}{\sqrt{K}}}\int_{\mathbb{R}} \mathcal{D}z e^{z(\lambda_l - i\frac{\hat{w}}{\sqrt{K}})} e^{-\frac{i\hat{u}}{\sqrt{K}}\mathrm{sign}[z]}. \tag{D.5}$$

We (abusively) dropped the dependency of λ_l on (w,ξ). Note the following identity:

$$\int_{\mathbb{R}} \mathcal{D}z\, e^{\alpha z + i\beta\,\mathrm{sign}(z)} = e^{\alpha^2/2}[\cos\beta + i\sin\beta\hat{H}(\alpha)], \tag{D.6}$$

with $\hat{H}(x) = \mathrm{erf}(x/\sqrt{2})$. Using eq. (D.6) in eq. (D.5) we obtain:

$$J(\lambda_l,\hat{w},\hat{u}) = e^{-\frac{1}{2K}\hat{w}^2}\left[\cos\left(\frac{\hat{u}}{\sqrt{K}}\right) - i\sin\left(\frac{\hat{u}}{\sqrt{K}}\right)\hat{H}\left(\lambda_l - i\frac{\hat{w}}{\sqrt{K}}\right)\right].$$

By our committee-symmetry assumption, we have $\lambda_l(w,\xi) = \lambda_{l,0}(\xi) + \lambda_1(w,\xi)/\sqrt{K}$, with $\lambda_{l,0}$ and λ_1 typically of order 1 when $K \to \infty$ and given by:

$$\begin{cases} \lambda_{l,0}(\xi) & \equiv \sqrt{\dfrac{q_d}{1-q_d}}\xi_l, \\ \lambda_1(w,\xi) & \equiv \left[\sqrt{\dfrac{1-q_a-q_d}{1-q_d}}-1\right]w + \left[\sqrt{\dfrac{q_a+q_d}{1-q_d}} - \sqrt{\dfrac{q_d}{1-q_d}}\right]\dfrac{\mathbf{1}_K^\intercal\xi}{\sqrt{K}}. \end{cases} \tag{D.7}$$

Expanding $J(\lambda_l,\hat{w},\hat{u})$ as $K \to \infty$, we obtain using the asymptotic development of $\mathrm{erf}(x)$:

$$J(\lambda_l,\hat{w},\hat{u}) = e^{-\frac{1}{2K}\hat{w}^2}\left\{1 - \frac{\hat{u}^2}{2K} - i\hat{H}[\lambda_{l,0}(\xi)]\frac{\hat{u}}{\sqrt{K}} - i\frac{\hat{u}[\lambda_1(w,\xi)-i\hat{w}]}{K}\sqrt{\frac{2}{\pi}}e^{-\frac{\lambda_{l,0}(\xi)^2}{2}} + \mathcal{O}(K^{-3/2})\right\}.$$

This yields:

$$\prod_{l=1}^{K} J[\lambda_l(w,\xi),\hat{w},\hat{u})] = e^{-\frac{1}{2}\hat{w}^2 - \frac{\hat{u}^2}{2} - i\hat{u}S_1 - i\sqrt{\frac{2}{\pi}}\hat{u}(\lambda_1 - i\hat{w})\Gamma_0 + \frac{1}{2}\hat{u}^2 S_2 + \mathcal{O}(K^{-1/2})}, \tag{D.8}$$

in which we defined the following quantities, that only depend on ξ (recall eq. (D.7))

$$w_\xi(\xi) \equiv \frac{1}{\sqrt{K}}\sum_{l=1}^K \xi_l, \qquad \Gamma_0(\xi) \equiv \frac{1}{K}\sum_{l=1}^K e^{-\frac{1}{2}\lambda_{l,0}(\xi)^2},$$
$$S_1(\xi) \equiv \frac{1}{\sqrt{K}}\sum_{l=1}^K \hat{H}(\lambda_{l,0}(\xi)), \qquad S_2(\xi) \equiv \frac{1}{K}\sum_{l=1}^K \hat{H}(\lambda_{l,0}(\xi))^2.$$

A detailed calculation actually shows that the previous expansion of eq. (D.8) is valid up to $\mathcal{O}(K^{-1})$, and not only $\mathcal{O}(K^{-1/2})$. Recall also eq. (D.4), in which one can now readily perform the integration over all variables $w, \hat{w}, u, \hat{u}$ to obtain (dropping the ξ dependency in $w_\xi, \Gamma_0, S_1, S_2$):

$$I_C(y,\xi) = H\Bigg[-y\frac{S_1 + \sqrt{\frac{2}{\pi}} w_\xi \Gamma_0 \frac{\sqrt{q_d+q_a}-\sqrt{q_d}}{\sqrt{1-q_d}}}{\sqrt{1 - S_2 - \frac{2}{\pi}\Gamma_0^2 \frac{q_a}{1-q_d}}}\Bigg] + \mathcal{O}(K^{-1}), \tag{D.9}$$

in which $H(x) \equiv \int_x^\infty \mathcal{D}z = [1 - \mathrm{erf}(x/\sqrt{2})]/2$. Note that all quantities $w_\xi, \Gamma_0, S_1, S_2$ only depend on ξ via its empirical measure: this remark is what will make the integration over $\xi \in \mathbb{R}^K$ tractable. We compute it in the following, using theoretical physics methods. We denote the quantity that appears in eq. (D.9) as a function of $w_\xi, \Gamma_0, S_1, S_2$:

$$G(y, w_\xi, \Gamma_0, S_1, S_2) \equiv H\Bigg[-y\frac{S_1 + \sqrt{\frac{2}{\pi}} w_\xi \Gamma_0 \frac{\sqrt{q_d+q_a}-\sqrt{q_d}}{\sqrt{1-q_d}}}{\sqrt{1 - S_2 - \frac{2}{\pi}\Gamma_0^2 \frac{q_a}{1-q_d}}}\Bigg].$$

Introducing once again delta functions and their Fourier transforms for $w_\xi, \Gamma_0, S_1, S_2$, we write, starting from eq. (D.9):

$$\begin{aligned} I_C &= \sum_{y=\pm 1}\int_{\mathbb{R}^K} \mathcal{D}\xi I_C(y,\xi)\ln I_C(y,\xi), \\ &= \sum_{y=\pm 1}\int \frac{\mathrm{d}w_\xi \mathrm{d}\hat{w}_\xi}{2\pi}\frac{\mathrm{d}\Gamma_0 \mathrm{d}\hat{\Gamma}_0}{2\pi}\frac{\mathrm{d}S_1 \mathrm{d}\hat{S}_1}{2\pi}\frac{\mathrm{d}S_2 \mathrm{d}\hat{S}_2}{2\pi} e^{iw\hat{w} + i\Gamma_0\hat{\Gamma}_0 + iS_1\hat{S}_1 + iS_2\hat{S}_2}\, G(y, w_\xi, \Gamma_0, S_1, S_2) \\ &\quad \times \ln G(y, w_\xi, \Gamma_0, S_1, S_2)\Big[\int_{\mathbb{R}^K} \mathcal{D}\xi e^{-i\hat{w}w_\xi(\xi) - i\hat{\Gamma}_0\Gamma_0(\xi) - i\hat{S}_1 S_1(\xi) - i\hat{S}_2 S_2(\xi)}\Big] + \mathcal{O}(K^{-1}). \end{aligned} \tag{D.10}$$

The integral over ξ in eq. (D.10) can be computed in the limit $K \to \infty$:

$$\begin{aligned} \Lambda &\equiv \int_{\mathbb{R}^K} \mathcal{D}\xi\, e^{-i\hat{w}w_\xi(\xi) - i\hat{\Gamma}_0\Gamma_0(\xi) - i\hat{S}_1 S_1(\xi) - i\hat{S}_2 S_2(\xi)} \\ &= \Bigg\{ \int_{\mathbb{R}} \mathcal{D}\xi \exp\Big[-i\frac{\hat{w}\xi}{\sqrt{K}} - i\frac{\hat{\Gamma}_0 e^{-\frac{q_d}{2(1-q_d)}\xi^2}}{K} - i\frac{\hat{S}_1 \hat{H}\Big[\sqrt{\frac{q_d}{1-q_d}}\xi\Big]}{\sqrt{K}} - i\frac{\hat{S}_2 \hat{H}\Big[\sqrt{\frac{q_d}{1-q_d}}\xi\Big]^2}{K}\Big]\Bigg\}^K \end{aligned}$$

The large K expansion of this expression yields

$$\begin{aligned} \Lambda = \exp\Bigg\{ &-\frac{1}{2}\hat{w}^2 - i\hat{\Gamma}\sqrt{1-q_d} - \hat{S}_1\hat{w}\mathbb{E}\Big[\xi\hat{H}\Big(\sqrt{\frac{q_d}{1-q_d}}\xi\Big)\Big] \\ &- \frac{\hat{S}_1^2 + i\hat{S}_2}{2}\mathbb{E}\Big[\hat{H}\Big(\sqrt{\frac{q_d}{1-q_d}}\xi\Big)^2\Big]\Bigg\} + \mathcal{O}(K^{-1}). \end{aligned}$$

The expectations are taken with respect to a real variable $\xi \sim \mathcal{N}(0,1)$. They are known properties of the error function:

$$\mathbb{E}\Big[\hat{H}\Big(\sqrt{\frac{q_d}{1-q_d}}\xi\Big)^2\Big] = \frac{2}{\pi}\arcsin q_d, \quad \text{and} \quad \mathbb{E}\Big[\xi\hat{H}\Big(\sqrt{\frac{q_d}{1-q_d}}\xi\Big)\Big] = \sqrt{\frac{2q_d}{\pi}}.$$

One can now compute the integrals over the "hat" variables in eq. (D.10). Denote $\Gamma_0^f \equiv \sqrt{\frac{2(1-q_d)}{\pi}}$, and $S_2^f \equiv \frac{2}{\pi}\arcsin q_d$. This yields:

$$\begin{aligned} I_C = \int_{\mathbb{R}^2} \mathcal{D}w\mathcal{D}S_1\, G\Big(y, w, \Gamma_0^f, \sqrt{\frac{2(\arcsin q_d - q_d)}{\pi}}S_1 + w\sqrt{\frac{2q_d}{\pi}}, S_2^f\Big) \\ \times \ln G\Big(y, w, \Gamma_0^f, \sqrt{\frac{2(\arcsin q_d - q_d)}{\pi}}S_1 + w\sqrt{\frac{2q_d}{\pi}}, S_2^f\Big). \end{aligned} \tag{D.11}$$

Note that

$$G\Big(y, w, \Gamma_0^f, \sqrt{\frac{2(\arcsin q_d - q_d)}{\pi}}S_1 + w\sqrt{\frac{2q_d}{\pi}}, S_2^f\Big) = H\Big[-y\sqrt{\frac{2}{\pi}}\frac{\sqrt{\arcsin q_d - q_d}S_1 + w\sqrt{q_d+q_a}}{\sqrt{1-\frac{2}{\pi}(q_a+\arcsin q_d)}}\Big].$$

Making the change of variable $S_1^{\text{new}} = S_1 + w\frac{\sqrt{q_d+q_a}}{\sqrt{\arcsin q_d - q_d}}$ in eq. (D.11), and defining $\gamma \equiv \frac{2}{\pi}(q_a + \arcsin q_d)$, one reaches:

$$I_C = \sum_{y=\pm 1}\int_{\mathbb{R}} \mathcal{D}x H\Big[yx\sqrt{\frac{\gamma}{1-\gamma}}\Big]\ln H\Big[yx\sqrt{\frac{\gamma}{1-\gamma}}\Big] + \mathcal{O}(K^{-1}).$$

The two values $y = \pm 1$ contribute in the same way, which finally yields:

$$I_C = 2\int_{\mathbb{R}} \mathcal{D}x H\Big[x\sqrt{\frac{\gamma}{1-\gamma}}\Big]\ln H\Big[x\sqrt{\frac{\gamma}{1-\gamma}}\Big] + \mathcal{O}(K^{-1}). \tag{D.12}$$

Note that the parameter γ is naturally bounded to the interval $[0,1]$ by the conditions $0 \le q_d \le 1$ and $0 \le q_a + q_d \le 1$.

D.2.2 Limit of the prior integral

The prior part I_P of the free entropy of eq. (B.6) is very easy to evaluate in the Gaussian prior setting. Recall that we consider a prior with covariance matrix $Q_0 = \mathrm{I}_K$. Performing the Gaussian integration in I_P yields:

$$I_P = \frac{K}{2}\hat{q}_d + \frac{1}{2}\hat{q}_a - \frac{K-1}{2}\ln(1+\hat{q}_d) - \frac{1}{2}\ln(1+\hat{q}_d+\hat{q}_a). \tag{D.13}$$

D.2.3 Limit of the State Evolution

From the definition of the free entropy in eq. (B.6) and the expansions for I_P and I_C obtained in (D.12) and (D.13), one obtains the fixed point equations after having extremized over $\hat{q}_d$ and $\hat{q}_a$ (recall that $\alpha \equiv m/n$):

$$\begin{cases} \partial_{q_a}[I_G(q_d,q_a) + \alpha I_C(q_d,q_a)] &= 0, \\ \partial_{q_d}[I_G(q_d,q_a) + \alpha I_C(q_d,q_a)] &= 0, \end{cases} \tag{D.14}$$

with I_G, I_c defined as:

$$I_G(q_d, q_a) \equiv \frac{1}{2}[q_a + Kq_d] + \frac{K-1}{2}\ln[1-q_d] + \frac{1}{2}\ln[1-q_a-q_d],$$
$$I_C(q_d, q_a) \equiv 2\int_{\mathbb{R}} \mathcal{D}x H\Big[x\sqrt{\frac{\gamma}{1-\gamma}}\Big]\ln H\Big[x\sqrt{\frac{\gamma}{1-\gamma}}\Big],$$

and recall that $\gamma \equiv \frac{2}{\pi}(q_a + \arcsin q_d)$. The fixed point equations (D.14) have different behavior depending on the scaling of α with the hidden layer size K. We detail them in the following paragraphs.

Regime $\alpha = \mathcal{O}_{K\to\infty}(K)$

In this regime (which in particular contains the case in which α stays of order 1 when $K \to \infty$), the fixed point equations (D.14) can be simplified as:

$$\begin{cases} q_d = 0, \\ q_a = 2\alpha(1-q_a)\frac{\partial \mathcal{I}_C}{\partial q_a}. \end{cases} \tag{D.15}$$

Regime $\alpha = \Theta_{K\to\infty}(K)$

In this regime, we naturally define $\widetilde{\alpha}K \equiv \alpha/K$, such that $\widetilde{\alpha}$ will remain of order 1. One can show that the solutions of the fixed point equations (D.14) must satisfy the following scaling: $q_a + q_d = 1 - \chi/K$, with $\chi \geq 0$ reaching a finite value when $K \to \infty$. The fixed point equations in terms of χ and q_d read:

$$\begin{cases} q_d = 2(1-q_d)\Big(\frac{1}{\sqrt{1-q_d^2}} - 1\Big)\tilde{\alpha}\frac{\partial \mathcal{I}_C}{\partial q_a}, \\ \chi^{-1} = 2\widetilde{\alpha}\frac{\partial \mathcal{I}_C}{\partial q_a}. \end{cases} \tag{D.16}$$

Note that the State Evolution (SE) computation of Figure 4.2 was performed by solving the fixed point equations (D.15) and (D.16) (depending on the regime of α).

Stability of the $q_d = 0$ non-specialized solution

It is easy to show that eq. (D.16) always admits what we call a *non-specialized solution*, i.e. a solution with $q_d = 0$. This solution stops to be globally optimal in term of the free entropy at a finite $\widetilde{\alpha}_{\text{spec}} \simeq 7.65$. However, one can show that this solution will remain *linearly* stable for every $\widetilde{\alpha}$. We can actually show that it is linearly stable in the much broader regime $\alpha = \mathcal{O}(K^2)$. Let us now justify this statement. Going back to the initial formulation of the fixed point equations (D.14), and adding the correct time indices to iterate them, one obtains:

$$q_d^{t+1} = \frac{F(q_d^t, q_a^t)}{1 + F(q_d^t, q_a^t)}, \tag{D.17}$$
$$q_a^{t+1} = \frac{G(q_d^t, q_a^t)}{(1 + F(q_d^t, q_a^t))(1 + F(q_d^t, q_a^t)G(q_d^t, q_a^t))}, \tag{D.18}$$

with F and G defined as:

$$F(q_d, q_a) \equiv \frac{2\alpha}{K-1}[\partial_{q_d} I_C - \partial_{q_a} I_C], \tag{D.19}$$
$$G(q_d, q_a) \equiv \frac{2\alpha K}{K-1}[\partial_{q_a} I_C - \frac{1}{K}\partial_{q_d} I_C]. \tag{D.20}$$

We focus on the behavior of (D.17) around $q_d = 0$. Given our previous expansion of I_C in the $K \to \infty$ limit (cf. eq. (D.12)), and eq. (D.19), one easily sees that for $\alpha = \mathcal{O}_{K\to\infty}(K^2)$, then $\partial_{q_d} F(q_d = 0) \to_{K\to\infty} 0$, which means the $q_d = 0$ solution always remains linearly stable. However, assume now that $\alpha = \Theta(K^2)$. Performing a similar calculation to the one described in Sec. D.2.1, one can show the following expansion:

$$I_C(q_d, q_a) = I_C^{(0)}(q_d, q_a) + \frac{1}{K} I_C^{(1)}(q_d, q_a) + \mathcal{O}\Big(\frac{1}{K^2}\Big).$$

The term of $\partial_{q_d} F(q_d = 0)$ arising from $I_C^{(1)}$ will thus have a possibly non-zero contribution in the $K \to \infty$ limit as soon as α/K^2 is no longer negligible, as seen from eq. (D.19).

To summarize, the non-specialized solution always remains linearly stable in the large K limit at least for $\alpha \ll K^2$. This implies that in this regime, Approximate Message Passing can not escape the non-specialized fixed point, as seen in Fig. 4.3. For α of order larger than K^2, one would have to explicitly compute $I_C^{(1)}$ in order to check that $\partial_{q_d} F(q_d = 0) \neq 0$, to show that the non-specialized solution is indeed linearly unstable. This verification of actual instability for $\alpha \sim K^2$ is not done in this thesis, and remains to be conducted.

D.2.4 The generalization error at large K

Recall the definition of the generalization error in eq. (D.1). Having this definition in mind, and recalling that this generalization error only depends on the asymptotic overlap, one can compute it at large K by applying the same techniques used to compute the channel integral I_C in Sec. D.2.1. One obtains after a tedious (but straightforward) calculation:

$$\epsilon_g^{\text{Bayes}} = \frac{1}{\pi} \arccos\Big[\frac{2}{\pi}(q_a + \arcsin q_d)\Big] + \mathcal{O}(K^{-1}). \tag{D.21}$$

This expression is the one used in the computation of the generalization error in the left panel of Fig. 4.3.

D.3 RMT analysis of the spiked matrix model

D.3.1 Proof of Lemma 5.6

Point (i) is trivial by definition of $S_k^{(r)}(\lambda)$ and the result of Lemma 5.5. We turn to points (ii) and (iii). Let us denote the following function:

$$T^{(2)}(s) \equiv s[\alpha(1+\alpha) - (1+2\alpha)(1 + s g_\nu^{-1}(s)) + (1 + s g_\nu^{-1}(s))^2].$$

By Lemma 5.5, we have $T^{(2)}(s) = S^{(2)}(g_\nu^{-1}(s))$, so $T^{(2)}(s) < 0$ for $s \in (s_{\text{edge}}, 0)$ by negativity of $S^{(2)}(\lambda)$. Therefore, point (ii) is equivalent to:

$$\forall s \in (s_{\text{edge}}, 0), \quad T^{(2)}(s) = -\alpha\Delta \Leftrightarrow s = g_\nu(1) \text{ and } \Delta \leq \Delta_c(\alpha), \tag{D.22}$$

while point (iii) means that for every $\Delta > \Delta_c(\alpha)$,

$$\forall s \in (s_{\text{edge}}, 0), \quad T^{(2)}(s) > -\alpha\Delta. \tag{D.23}$$

The condition $s > s_{\text{edge}}$ arises naturally as the counterpart of $z \geq \lambda_{\max}$. Recall that by Theorem 5.2, we have $\lambda_{\max} \leq 1$ for all Δ. As $g_\nu^{-1}(s)$ is here completely explicit by eq. (5.26), and

recalling the form of ρ_Δ in eq. (5.23), it is easy to show by an explicit computation that:

$$\forall s \neq -1, \quad T^{(2)}(s) = -\alpha\Delta + \alpha[g_\nu^{-1}(s) - 1]\frac{s - \Delta - 2s\Delta + \sqrt{s^2 - 2s(1+s)\Delta + \Delta^2}}{2(1+s)},$$

$$T^{(2)}(-1) = \begin{cases} -\alpha(1+\alpha) & \text{if } \Delta \geq 1, \\ -\alpha\Delta(1+\alpha\Delta) & \text{if } \Delta \leq 1. \end{cases}$$

It is then easy to see that the only possible solution to $T(s) = -\alpha\Delta$ with $s \in (s_{\text{edge}}, 0)$ is $s = g_\nu(1)$, if $g_\nu(1) \neq -1$. However, by Lemma 5.4, for any $\Delta > \Delta_c(\alpha)$ we have $s_{\text{edge}} < -1$. Moreover, in this case, one computes very easily (all expressions are explicit) $g_\nu^{-1}(-1) = 1$. Given the identity above, there is therefore no solution to $T^{(2)}(s) = -\alpha\Delta$ in $(s_{\text{edge}}, 0)$. By continuity of $T^{(2)}(s)$, and since $\lim_{s\to 0} T^{(2)}(s) = 0$, this implies $T^{(2)}(s) > -\alpha\Delta$ for $s \in (s_{\text{edge}}, 0)$, which proves point (iii).

Assume now $\Delta \leq \Delta_c(\alpha)$. Note that the case $\Delta = \Delta_c(\alpha)$ is easy, as $s_{\text{edge}} = -1$ is the unique solution to $T^{(2)}(s) = -\alpha(1+\alpha)$. For $\Delta < \Delta_c(\alpha)$, by Lemma 5.4 we obtain $-1 < s_{\text{edge}}$. In particular, $g_\nu(1) > s_{\text{edge}} > -1$, and we thus have that $s = g_\nu(1)$ is a solution (and the only one) to $T^{(2)}(s) = -\alpha\Delta$ by the identity shown above. This shows (ii) and ends the proof of Lemma 5.6.

D.3.2 Proof of correlation of the leading eigenvector

We now turn to the study of the leading eigenvector in the claim of Theorem 5.3. Let $\tilde{\mathbf{v}}$ be an eigenvector associated with the largest eigenvalue λ_1, normalized such that $\|\tilde{\mathbf{v}}\|^2 = p$. We have:

$$(\lambda_1 \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})\tilde{\mathbf{v}} = \frac{1}{\Delta}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\frac{\mathbf{v}^\intercal\tilde{\mathbf{v}}}{p}\mathbf{v}. \tag{D.24}$$

By normalization of $\tilde{\mathbf{v}}$, we obtain:

$$\tilde{\mathbf{v}} = \sqrt{p}\frac{(\lambda_1 \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{v}}{\sqrt{\mathbf{v}^\intercal\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{A}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{v}}}, \qquad \text{with} \qquad \mathbf{A} \equiv (\lambda_1 \mathrm{I}_p - (\mathbf{\Gamma}_p^{(0)})^\intercal)^{-1}(\lambda_1 \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}.$$

Therefore:

$$\frac{1}{p^2}|\tilde{\mathbf{v}}^T\mathbf{v}|^2 = \frac{1}{p}\frac{[\mathbf{v}^\intercal(\lambda_1 \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{v}]^2}{\mathbf{v}^\intercal\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{A}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{v}}.$$

Using $\mathbf{v} = \mathbf{W}\mathbf{z}/\sqrt{k}$ and the concentration of $\mathbf{z}^\intercal\mathbf{M}\mathbf{z}/k$ on $\operatorname{Tr}\mathbf{M}/k$, we reach as $p, k \to \infty$:

$$\frac{1}{p^2}|\tilde{\mathbf{v}}^T\mathbf{v}|^2 \simeq \frac{\left[\frac{1}{p}\operatorname{Tr}\{(\lambda_1 \mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}(\frac{\mathbf{W}\mathbf{W}^\intercal}{k})^2\}\right]^2}{\frac{1}{p}\operatorname{Tr}\{\mathbf{A}(\frac{\mathbf{W}\mathbf{W}^\intercal}{k})^3\}}. \tag{D.25}$$

The numerator is equal to $[\alpha^{-1}S_k^{(2)}(\lambda_1)]^2$, (recall that $S_k^{(r)}$ is defined in Lemma 5.5). Let us compute the denominator. Recall that we can write $\mathbf{\Gamma}_p^{(0)} = \mathbf{W}\mathbf{W}^\intercal\mathbf{M}/k$, with a symmetric

matrix $\mathbf{M}$ that is independent of $\mathbf{W}$. For any z large enough, we can expand:

$$\begin{aligned}
&\mathrm{Tr}\Big\{(z\mathrm{I}_p - (\mathbf{\Gamma}_p^{(0)})^\intercal)^{-1}(z\mathrm{I}_p - \mathbf{\Gamma}_p^{(0)})^{-1}\Big(\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\Big)^3\Big\}\\
&= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty} z^{-a-b-2}\,\mathrm{Tr}\Big\{\Big(\mathbf{M}\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\Big)^a\Big(\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\mathbf{M})\Big)^b\Big(\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\Big)^3\Big\},\\
&\overset{(a)}{=} \sum_{a=0}^{\infty}\sum_{b=0}^{\infty} z^{-a-b-2}\,\mathrm{Tr}\Big\{\Big(\frac{\mathbf{W}^\intercal\mathbf{M}\mathbf{W}}{k}\Big)^a\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big(\frac{\mathbf{W}^\intercal\mathbf{M}\mathbf{W}}{k}\Big)^b\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^2\Big\},\\
&= \mathrm{Tr}\Big\{(z\mathrm{I}_k - \mathbf{\Gamma}_k^{(0)})^{-1}\frac{\mathbf{W}^\intercal\mathbf{W}}{k}(z\mathrm{I}_k - \mathbf{\Gamma}_k^{(0)})^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^2\Big\} = kS_k^{(1,2)}(z),
\end{aligned}$$

where in (a) we used the cyclicity of the trace. Given the last statement of Theorem 5.2, we know $\liminf_{p\to\infty}\lambda_1 \geq \lambda_{\max}$, so we can use the above calculation to write from eq. (D.25):

$$\epsilon(\Delta) = \lim_{\lambda\downarrow\lambda_1}\lim_{k\to\infty}\frac{1}{\alpha}\frac{[S_k^{(2)}(\lambda)]^2}{S_k^{(1,2)}(\lambda)}. \tag{D.26}$$

As in the eigenvalue transition proof, to make this fully rigorous one would need to use more precisely concentration results, and would follow exactly the lines of [BGN11]. We now use the (already proven) transition of the leading eigenvalue, that gives us the value of λ_1.

- For $\Delta < \Delta_c(\alpha)$, we know that λ_1 converges almost surely to 1. Consequently:

$$\epsilon(\Delta) = \frac{1}{\alpha}\frac{[S^{(2)}(1)]^2}{S^{(1,2)}(1)}.$$

  By Lemma 5.6, we know that $S^{(2)}(1) = -\alpha\Delta$. Moreover $\lambda_{\max} < 1$ by Theorem 5.2. This implies that $S^{(1,2)}(1)$ is finite and strictly positive, from its very definition. Therefore for every $\Delta < \Delta_c(\alpha)$, $\epsilon(\Delta) > 0$.

- For $\Delta = \Delta_c(\alpha)$, we have $\lambda_{\max} = 1$ and $\lim_{\lambda\to 1} S^{(2)}(\lambda) = -\alpha\Delta$ as we have shown. For every r, q, let us define the functions $T^{(r)}$ and $T^{(r,q)}$ by $S^{(r)}(\lambda) = T^{(r)}[g_\nu(\lambda)]$ and $S^{(r,q)}(\lambda) = T^{(r,q)}[g_\nu(\lambda)]$. By Lemma 5.5 and the chain rule, we have for all $s(s_{\mathrm{edge}}, 0)$:

$$\begin{aligned}
T^{(1,2)}(s) = sT^{(3)}(s) - [1 + sg_\nu^{-1}(s)]\Big[T^{(1,1)}(s) + (1+\alpha)\frac{\partial_s T^{(1)}(s)}{\partial_s g_\nu^{-1}(s)}\Big]\\
+ \alpha s[(1+\alpha)s + T^{(1)}(s) + T^{(2)}(s)]\int\frac{\rho_\Delta(\mathrm{d}t)t}{(1+ts)^2}\Big[t\frac{\partial_s T^{(1)}(s)}{\partial_s g_\nu^{-1}(s)} - s\Big].
\end{aligned} \tag{D.27}$$

  Recall that $g_\nu^{-1}(s)$ is explicit by eq. (5.26) and $s_{\mathrm{edge}} = \lim_{\lambda\downarrow\lambda_{\max}} g_\nu(\lambda)$. It moreover satisfies (cf. Theorem 5.2) $\partial_s g_\nu^{-1}(s_{\mathrm{edge}}) = 0$. For $\Delta = \Delta_c(\alpha)$, by Lemma 5.4 we have $g_\nu(1) = -1 = s_{\mathrm{edge}}$. It is then only trivial algebra to verify from eq. (D.27) and the remaining relations of Lemma 5.5 that $T^{(1,2)}(-1) = +\infty$, which implies $\epsilon(\Delta_c(\alpha)) = 0$.

- We investigate here the $\Delta \to 0$ limit. In this limit, we know from eq. (D.26) and the analysis in the case $\Delta < \Delta_c(\alpha)$ above that

$$\lim_{\Delta\to 0}\epsilon(\Delta) = \lim_{\Delta\to 0}\frac{\alpha\Delta^2}{S^{(1,2)}(1)}.$$

It is again heavy but straightforward algebra to verify from eq. (D.27) and the remaining relations of Lemma 5.5 that as $\Delta \to 0$ and for any $s \in (s_{\text{edge}}, 0)$:

$$T^{(1,2)}(s) = \alpha\Delta^2 + \mathcal{O}(\Delta^3).$$

This yields $\lim_{\Delta\to 0} \epsilon(\Delta) = 1$.

- Finally, we consider $\Delta > \Delta_c(\alpha)$. By eq. (D.26) and item (*iii*) of Lemma 5.6, to obtain $\epsilon(\Delta) = 0$ we only need to prove that $\lim_{\lambda\to\lambda_{\max}} S^{(1,2)}(\lambda) = +\infty$. Equivalently, we must show $\lim_{s\to s_{\text{edge}}} T^{(1,2)}(s) = +\infty$. Recall that $\partial_s g_\nu^{-1}(s_{\text{edge}}) = 0$ and that since s_{edge} is finite, all $T^{(r)}(s_{\text{edge}})$ for $r = 0, 1, 2, 3$ are finite as well by Lemma 5.5. It thus only remains to check that $\lim_{s\to s_{\text{edge}}} T^{(1,2)}(s)\partial_s\, g_\nu^{-1}(s) > 0$. This would imply that $\lim_{s\to s_{\text{edge}}} T^{(1,2)}(s) = +\infty$. We put this statement as a lemma, actually stronger than what we need:

Lemma D.1 (*Lower bound on* $T^{(1,2)}$)

For every $\alpha > 0$ and $\Delta > 1$, we have $\liminf_{s\to s_{\text{edge}}} T^{(1,2)}(s)\partial_s g_\nu^{-1}(s) > 0$.

We prove this for every $\Delta > 1$, while only the case $\Delta > \Delta_c = 1 + \alpha$ is needed in our analysis. As already argued, this lemma ends the proof of the eigenvector correlation in Theorem 5.3.

Proof of Lemma D.1 – The idea is to lower bound $S^{(1,2)}(\lambda)$ by $\partial_\lambda g_\nu(\lambda)$, for every $\lambda > \lambda_{\max}$. We separate three cases:

- First, assume $\alpha > 1$. Then $\mathbf{W}^\intercal\mathbf{W}/k$ has full rank. In particular, by the classical results of [MP67], its lowest eigenvalue, denoted $\zeta_{\min}$, converges almost surely to $(1-\alpha^{-1/2})^2$. Moreover, for any two symmetric positive square matrices $\mathbf{A}$ and $\mathbf{B}$, we know that $\operatorname{Tr}[\mathbf{AB}] \geq 0$[2]. This implies immediately that if a_0 is the smallest eigenvalue of $\mathbf{A}$, then $\operatorname{Tr}[\mathbf{AB}] \geq a_0 \operatorname{Tr}[\mathbf{B}]$, as $\mathbf{A} - a_0\mathrm{I}$ is positive. We can use this to write, for any $\lambda > \lambda_{\max}$:

$$\begin{aligned} S_k^{(1,2)}(\lambda) &= \frac{1}{k}\operatorname{Tr}\Big[(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)^2\Big], \\ &\geq \zeta_{\min}^2 \frac{1}{k}\operatorname{Tr}\Big[(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}\Big(\frac{\mathbf{W}^\intercal\mathbf{W}}{k}\Big)(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-1}], \\ &\geq \zeta_{\min}^3 \frac{1}{k}\operatorname{Tr}[(\mathbf{\Gamma}_k^{(0)} - \lambda \mathrm{I}_k)^{-2}]. \end{aligned}$$

  Taking the limit $k \to \infty$ in this last inequality, we obtain:

$$S^{(1,2)}(\lambda) \geq (1 - \alpha^{-1/2})^6 \partial_\lambda g_\nu(\lambda).$$

  Taking the limit $\lambda \to \lambda_{\max}$ (or equivalently $s \to s_{\text{edge}}$) yields the sought result:

$$\liminf_{s\to s_{\text{edge}}} T^{(1,2)}(s)\partial_s g_\nu^{-1}(s) \geq (1 - \alpha^{-1/2})^6 > 0. \tag{D.28}$$

- Assume now $\alpha < 1$. We do the same reasoning, as $\mathbf{W}\mathbf{W}^\intercal/k$ has now full rank, and its smallest eigenvalue, also denoted $\zeta_{\min}$, converges a.s. as $k \to \infty$ to $(1 - \sqrt{\alpha})^2$. We know that we can rewrite $S_k^{(1,2)}(\lambda)$ as the trace of a $p \times p$ matrix:

$$\begin{aligned} S_k^{(1,2)}(\lambda) &= \frac{1}{k}\operatorname{Tr}[((\mathbf{\Gamma}_p^{(0)})^\intercal - \lambda \mathrm{I}_p)^{-1}(\mathbf{\Gamma}_p^{(0)} - \lambda \mathrm{I}_p)^{-1}\Big(\frac{\mathbf{W}\mathbf{W}^\intercal}{k}\Big)^3], \\ &\geq \zeta_{\min}^3 \frac{1}{k}\operatorname{Tr}[((\mathbf{\Gamma}_p^{(0)})^\intercal - \lambda \mathrm{I}_p)^{-1}(\mathbf{\Gamma}_p^{(0)} - \lambda \mathrm{I}_p)^{-1}] \geq \zeta_{\min}^3 \frac{1}{k}\operatorname{Tr}[(\mathbf{\Gamma}_p^{(0)} - \lambda \mathrm{I}_p)^{-2}], \end{aligned}$$

[2] Indeed, there exists a positive square root of $\mathbf{A}$, and $\operatorname{Tr}[\mathbf{AB}] = \operatorname{Tr}[\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}] \geq 0$.

in which the last inequality comes from $\mathrm{Tr}\,[\mathbf{A}\mathbf{A}^\intercal] \geq \mathrm{Tr}\,[\mathbf{A}^2]$ for any positive square matrix $\mathbf{A}$. Once again, taking the limit $k \to \infty$, and then the limit $\lambda \to \lambda_{\max}$, this yields

$$\liminf_{s \to s_{\mathrm{edge}}} T^{(1,2)}(s) \partial_s g_\nu^{-1}(s) \geq (1 - \alpha^{1/2})^6 > 0. \tag{D.29}$$

- Finally, we treat the $\alpha = 1$ case. In this case, we can not use easy bounds as in the two previous cases, since the support of the Marchenko-Pastur distribution touches 0. However, recall that everything is explicit here : ρ_Δ is given by eq. (5.23), $g_\nu^{-1}(s)$ is given by eq. (5.26) and Lemma 5.5 gives all the $T^{(r)}$ and $T^{(r,q)}$ in terms of g_ν^{-1} and ρ_Δ. We can moreover use what we proved in Theorem 5.2:

$$\partial_s g_\nu^{-1}(s_{\mathrm{edge}}) = \frac{1}{s^2} - \alpha \int \rho_\Delta(\mathrm{d}t) \frac{t^2}{(1 + t s_{\mathrm{edge}})^2} = 0.$$

This can be used to simplify the term $\partial_s T^{(1)}(s)$ and the term $\int \rho_\Delta(\mathrm{d}t)[t/(1+ts)]^2$. Some heavy but straightforward algebra yields from these relations that the following limit is finite, and is given by:

$$\lim_{s \to s_{\mathrm{edge}}} T^{(1,2)}(s)\, \partial_s g_\nu^{-1}(s) = h(s_{\mathrm{edge}}),$$

with

$$h(s) = \frac{h_1(s)^2 \times h_2(s)}{4s^6}, \text{ and } \begin{cases} h_1(s) &= -\Delta + \sqrt{\Delta^2 + s^2 - 2\Delta(2s+1)s} + s, \\ h_2(s) &= 3\Delta - 3\sqrt{\Delta^2 + s^2 - 2\Delta(2s+1)s} + s(4s-3). \end{cases}$$

It is then very simple algebra (solving quadratic equations and using $\Delta > 1$) to see that there is no real negative solution to $h(s) = 0$, and that $h(s) > 0$ for all $s \in (-\infty, 0)$. This implies that $h(s_{\mathrm{edge}}) > 0$, which ends the proof.

□

D.4 State evolution of spectral methods with generative prior

As we have already mentioned in Section 5.2.4, one of the greatest virtues of AMP is being able to track its asymptotic performance through a set of simple scalar state evolution equations. Note that for the noiseless linear channel $P_{\mathrm{out}}(v|x) = \delta(v - x)$, Algorithm 4 is already linear! As a consequence, the state evolution equations associated to the spectral method are simply dictated by the set of AMP state evolution equations eq. (5.18).

However, it is worth stressing that as LAMP returns a normalized estimator, the LAMP MSE is not given by the AMP mean squared error. We now compute the overlaps and mean squared error performed by this spectral algorithm.

MSE achieved by LAMP – Recall that m_v and q_v are the parameters defined in eq. (5.16), respectively measuring the overlaps between the ground truth $\mathbf{v}^\star$ and the estimator $\hat{\mathbf{v}}$, and the norm of the estimator. In the general case, the MSE of eq. (5.7) becomes:

$$\mathrm{MSE}_v = \rho_v + \mathbb{E}_{\mathbf{v}^\star} \lim_{p \to \infty} \frac{1}{p} \|\hat{\mathbf{v}}\|_2^2 - 2\mathbb{E}_{\mathbf{v}^\star} \lim_{p \to \infty} \frac{1}{p} \hat{\mathbf{v}}^\intercal \mathbf{v}^\star = \rho_v + q_v - 2m_v\,, \tag{D.30}$$

However the LAMP spectral method computes the normalized leading eigenvector of the structured matrix $\mathbf{\Gamma}_p^{vv}$. Hence the norm of the LAMP estimator is $\|\hat{\mathbf{v}}\|^2_{\text{LAMP}} = q_{v,\text{LAMP}} = 1$, while the Bayes-optimal AMP estimator is not normalized, and satisfies $\|\hat{\mathbf{v}}\|^2_{\text{AMP}} = q^\star_{v,\text{AMP}} = m^\star_{v,\text{AMP}} \neq 1$, with $q^\star_{v\text{AMP}}$ solution of eq. (5.18). As the non-normalized LAMP estimator follows the AMP state evolution in the *linear case*, the overlap with the ground truth is thus given by:

$$m_{v,\text{LAMP}} \equiv \mathbb{E}_{\mathbf{v}^\star} \lim_{p\to\infty} \frac{1}{p} \hat{\mathbf{v}}^\intercal_{\text{LAMP}} \mathbf{v}^\star = \mathbb{E}_{\mathbf{v}^\star} \lim_{p\to\infty} \frac{1}{p} \Big(\frac{\hat{\mathbf{v}}_{\text{AMP}}}{\|\hat{\mathbf{v}}\|_{\text{AMP}}}\Big)^\intercal \mathbf{v}^\star = \frac{m^\star_{v,\text{AMP}}}{(q^\star_{v,\text{AMP}})^{1/2}} = (m^\star_{v,\text{AMP}})^{1/2}\,.$$

Therefore the mean squared error performed by the LAMP method is easily obtained from the optimal overlap reached by the AMP algorithm and yields

$$\text{MSE}_{v,\text{LAMP}} = \rho_v + 1 - 2(q^\star_{v,\text{AMP}})^{1/2}\,. \tag{D.31}$$

MSE achieved by PCA – The respective result for PCA can be obtained from the observation that for the linear case, the $\alpha = 0$ LAMP operator reduces exactly to the matrix $\mathbf{Y}$. Therefore we can simply state that the mean squared error performed by PCA is computed using the optimal overlap reached by AMP at $\alpha = 0$:

$$\text{MSE}_{v,\text{PCA}} = \rho_v + 1 - 2(q^\star_{v,\text{AMP}}|_{\alpha=0})^{1/2}\,. \tag{D.32}$$

Comparing AMP with spectral methods – In order to fairly compare PCA, LAMP and AMP in Fig. 5.4, instead of showing the MSE corresponding to the *normalized* PCA and LAMP estimators (i.e. eqs. (D.31) and (D.32)), we rescale these spectral estimators by the optimal normalization $(q^\star_{v,\text{AMP}})^{1/2}$ (obtained from AMP). This is the convention used in Fig. 5.4, both in the linear and non-linear cases.

D.5 Derivation of thresholds in phase retrieval

D.5.1 Weak recovery

We detail here the derivation of the algorithmic weak-recovery threshold $\alpha_{\text{WR,Algo}}$. As discussed in Section 6.3.1, the weak-recovery threshold can be identified as the sample complexity for which the trivial fixed point $q_x = q_z = \hat{q}_x = \hat{q}_z = \gamma_x = \gamma_z = 0$ of the state evolution equations becomes linearly unstable. We repeat here the detailed state evolution for convenience:

$$\begin{cases}
q_x = \displaystyle\int_{\mathbb{K}} \mathcal{D}_\beta \xi \frac{\left|\int_{\mathbb{K}} P_0(\mathrm{d}x)\; x\; e^{-\frac{\beta}{2}\hat{q}_x|x|^2 + \beta\sqrt{\hat{q}_x}x\cdot\xi}\right|^2}{\int_{\mathbb{K}} P_0(\mathrm{d}x)\; e^{-\frac{\beta}{2}\hat{q}_x|x|^2 + \beta\sqrt{\hat{q}_x}x\cdot\xi}}, \\
q_z = \dfrac{1}{\hat{Q}_z + \hat{q}_z}\Big[\dfrac{\hat{q}_z}{\hat{Q}_z} + \displaystyle\int \mathrm{d}y\; \mathcal{D}_\beta \xi \frac{\left|\int \mathcal{D}_\beta z\; z\; P_{\text{out}}\Big(y\Big|\frac{z}{\sqrt{\hat{Q}_z+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi\Big)\right|^2}{\int \mathcal{D}_\beta z P_{\text{out}}\Big(y\Big|\frac{z}{\sqrt{\hat{Q}_z+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi\Big)}\Big], \\
\hat{q}_x = \dfrac{q_x}{\rho(\rho - q_x)} - \gamma_x, \\
\hat{q}_z = \dfrac{q_z}{Q_z(Q_z - q_z)} - \gamma_z, \\
\rho - q_x = \Big\langle \dfrac{1}{\rho^{-1} + \gamma_x + \lambda\gamma_z} \Big\rangle_\nu, \\
\alpha(Q_z - q_z) = \Big\langle \dfrac{\lambda}{\rho^{-1} + \gamma_x + \lambda\gamma_z} \Big\rangle_\nu.
\end{cases} \tag{D.33}$$

Letting $q_x = q_z = \hat{q}_x = \hat{q}_z = \gamma_x = \gamma_z = 0$, it is clear that the equations are satisfied if the signal distribution P_0 and the likelihood P_{out} satisfy the symmetry conditions of Def. 6.1. Assuming these conditions hold, we are interested in studying the linear stability of this local maximum. Recalling that $Q_z = \rho\langle\lambda\rangle_\nu/\alpha$, the first, third and fourth equations of (D.33) can be linearized:

$$\delta q_x = \rho^2 \delta\hat{q}_x, \qquad \delta\hat{q}_x = \frac{\delta q_x}{\rho^2} - \delta\gamma_x, \qquad \delta\hat{q}_z = \frac{\alpha^2\delta q_z}{\rho^2\langle\lambda\rangle_\nu^2} - \delta\gamma_z. \tag{D.34}$$

The second equation in (D.33) can be linearized to give:

$$\delta q_z = \frac{\rho^2\langle\lambda\rangle_\nu^2}{\alpha^2}\delta\hat{q}_z\Big(1 + \int_{\mathbb{R}} \mathrm{d}y \frac{\Big|\int_{\mathbb{K}} \mathcal{D}_\beta z\ (|z|^2-1)\ P_{\text{out}}(y|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha}}z)\Big|^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z\ P_{\text{out}}(y|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha}}z)}\Big). \tag{D.35}$$

Finally, it remains to compute the infinitesimal variation for $\delta\gamma_x, \delta\gamma_z$:

$$\begin{cases} \delta\gamma_x &= \dfrac{\langle\lambda^2\rangle_\nu}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_x - \dfrac{\alpha\langle\lambda\rangle_\nu}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_z, \\ \delta\gamma_z &= -\dfrac{\langle\lambda\rangle_\nu}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_x + \dfrac{\alpha}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_z. \end{cases} \tag{D.36}$$

Combining eqs. (D.34),(D.35),(D.36), we can simplify the system to a closed set equations over only $(\delta q_x, \delta\hat{q}_x, \delta q_z, \delta\hat{q}_z)$. Given the usual heuristics of the replica method and its link with message-passing algorithms [ZK16, TK20], one can easily check that the following time iteration of these equations corresponds to the state evolution of the G-VAMP algorithm:

$$\begin{cases} \delta q_x^{t+1} &= \rho^2\delta\hat{q}_x^t, \\ \delta q_z^{t+1} &= \dfrac{\rho^2\langle\lambda\rangle_\nu^2}{\alpha^2}\delta\hat{q}_z^t\Big(1 + \displaystyle\int_{\mathbb{R}} \mathrm{d}y \frac{\Big|\int_{\mathbb{K}} \mathcal{D}_\beta z\ (|z|^2-1)\ P_{\text{out}}(y|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha}}z)\Big|^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z\ P_{\text{out}}(y|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha}}z)}\Big), \\ \delta\hat{q}_x{}^t &= -\dfrac{\langle\lambda\rangle_\nu^2}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_x^t + \dfrac{\alpha\langle\lambda\rangle_\nu}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_z^t, \\ \delta\hat{q}_z{}^t &= \dfrac{\langle\lambda\rangle_\nu}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}\delta q_x^t + [\dfrac{\alpha^2}{\rho^2\langle\lambda\rangle_\nu^2} - \dfrac{\alpha}{\rho^2[\langle\lambda^2\rangle_\nu - \langle\lambda\rangle_\nu^2]}]\delta q_z^t. \end{cases} \tag{D.37}$$

From these equations, one can easily see that a linear instability of the trivial fixed points appears at $\alpha = \alpha_{\text{WR,Algo}}$ satisfying eq. (6.13). Indeed at $\alpha = \alpha_{\text{WR,Algo}}$, the modulus of all the eigenvalues of the size-4 matrix of the linear system (D.37) cross 1.

D.5.2 Perfect recovery

In this section, we assume a Gaussian standard prior $P_0 = \mathcal{N}_\beta(0,1)$ and a noiseless phase retrieval channel, and we show that information-theoretic full recovery is achieved exactly at $\alpha = \alpha_{\text{FR,IT}} \equiv \beta(1-\nu(\{0\}))$. We can assume without loss of generality that $\langle\lambda\rangle_\nu = \alpha$, as this amounts to a simple rescaling of $\mathbf{\Phi}$, irrelevant under the noiseless channel. This implies in particular that $Q_z = \hat{Q}_z = 1$.

The state evolution equations

Since we assumed a Gaussian prior we have, with $P_{\text{out}}(y|z) = \delta(y - |z|^2)$:

$$\begin{cases} q_z = \dfrac{1}{1+\hat{q}_z}\Big[\hat{q}_z + \displaystyle\int \mathrm{d}y \int_{\mathbb{K}} \mathcal{D}_\beta \xi \frac{\Big|\int_{\mathbb{K}} \mathcal{D}_\beta z \; z \; P_{\text{out}}\Big(y\Big|\frac{z}{\sqrt{1+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{1+\hat{q}_z}}\xi\Big)\Big|^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z P_{\text{out}}\Big(y\Big|\frac{z}{\sqrt{1+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{1+\hat{q}_z}}\xi\Big)}\Big], & \text{(D.38a)} \\ \hat{q}_x = \dfrac{q_x}{1-q_x}, & \text{(D.38b)} \\ \hat{q}_z = \dfrac{q_z}{1-q_z} - \gamma_z, & \text{(D.38c)} \\ q_x = \alpha\gamma_z(1-q_z), & \text{(D.38d)} \\ \alpha(1-q_z) = \Big\langle \dfrac{\lambda}{1+\lambda\gamma_z} \Big\rangle_\nu. & \text{(D.38e)} \end{cases}$$

Comparing these equations to Conjecture 6.1, one can see that in particular we have $\gamma_x = 0$, a straightforward consequence of the Gaussian prior.

Noisy phase retrieval with small variance

We wish to show that the free entropy of the full recovery solution is the global maximum of the free entropy potential for $\alpha > \alpha_{\text{IT}}$, while it is never the case for $\alpha < \alpha_{\text{IT}}$. However, under a noiseless channel, the free entropy potential might diverge in this point, which indicates towards a regularization procedure. Therefore we consider a noisy Gaussian channel with noise $\Delta > 0$, i.e. $P_{\text{out}}(y|z) = (2\pi\Delta)^{-1/2}\exp\{-(y-|z|^2)^2/(2\Delta)\}$. We will compute the limit, as $\Delta \downarrow 0$, of the free entropy of the "almost perfect" recovery fixed point. We look for a solution close to the point which corresponds to the best possible recovery, that is $q_z = 1$ and $q_x = 1-\nu(\{0\})$. Indeed it is easy to see that $q_x \leq 1-\nu(\{0\})$ since $\mathrm{rk}[\mathbf{\Phi}^\dagger\mathbf{\Phi}] \sim n(1-\nu(\{0\}))$. We are thus looking for a fixed point of the state evolution equations (D.38) that satisfies:

$$\begin{cases} q_x = 1-\nu(\{0\}) + \mathcal{O}_\Delta(1), & q_z = 1 + \mathcal{O}_\Delta(1), \\ \hat{q}_x^{-1} = \nu(\{0\})/(1-\nu(\{0\})) + \mathcal{O}_\Delta(1), & \hat{q}_z^{-1} = \mathcal{O}_\Delta(1). \end{cases} \tag{D.39}$$

Let us now precise the asymptotics of these quantities as $\Delta \downarrow 0$. By eq. (D.38d), we find easily:

$$\gamma_z \sim \frac{1-\nu(\{0\})}{\alpha(1-q_z)}. \tag{D.40}$$

Then from eq. (D.38c), we also have:

$$\hat{q}_z \sim \frac{\alpha - 1 + \nu(\{0\})}{\alpha(1-q_z)}. \tag{D.41}$$

Note that if $\alpha \leq 1$, then necessarily $\nu(\{0\}) \geq 1-\alpha$, so that the quantity in the numerator is always positive. We now turn to eq. (D.38a). We assume the scaling $\hat{q}_z^{-1} = c\Delta + \mathcal{O}_\Delta(\Delta)$. We

have by Gaussian integration by parts and using the specific form of P_{out}:

$$\int \mathrm{d}y \mathcal{D}_\beta \xi \frac{\left| \int \mathcal{D}_\beta z \; z \; P_{\text{out}}\left(y \middle| \frac{z}{\sqrt{1+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{1+\hat{q}_z}}\xi\right)\right|^2}{\int \mathcal{D}_\beta z P_{\text{out}}\left(y \middle| \frac{z}{\sqrt{1+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{1+\hat{q}_z}}\xi\right)}$$
$$= \frac{1}{(1+\hat{q}_z)} \int \mathrm{d}y \mathcal{D}_\beta \xi \frac{\left| \int \mathcal{D}_\beta z \; P'_{\text{out}}\left(y \middle| \frac{z}{\sqrt{1+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{1+\hat{q}_z}}\xi\right)\right|^2}{\int \mathcal{D}_\beta z P_{\text{out}}\left(y \middle| \frac{z}{\sqrt{1+\hat{q}_z}} + \sqrt{\frac{\hat{q}_z}{1+\hat{q}_z}}\xi\right)} \sim \frac{4}{\Delta(1+\hat{q}_z)} \sim 4c.$$

Note that in the complex case the derivative P'_{out} is here taken over z considered as an element of $\mathbb{R}^2$: it is not the Wirtinger derivative of complex analysis. This yields $1-q_z = \Delta c(1-4c)+\mathcal{O}_\Delta(1)$. Combining this result with eq. (D.41), we have

$$c(1-4c) = c\Big[\frac{\alpha - 1 + \nu(\{0\})}{\alpha}\Big].$$

This implies $c = (1-\nu(\{0\}))/(4\alpha)$, and we finally obtain the leading order asymptotics of $q_z, \hat{q}_z, \gamma_z$ as $\Delta \downarrow 0$:

$$\begin{cases} \hat{q}_z &= \dfrac{4\alpha}{(1-\nu(\{0\}))\Delta} + \mathcal{O}_\Delta(\Delta^{-1}), \\ 1-q_z &= \dfrac{(1-\nu(\{0\})(\alpha - 1 + \nu(\{0\})))}{4\alpha^2}\Delta + \mathcal{O}_\Delta(\Delta), \\ \gamma_z &= \dfrac{4\alpha}{\Delta(\alpha - 1 + \nu(\{0\}))} + \mathcal{O}_\Delta(\Delta^{-1}). \end{cases} \tag{D.42}$$

We then compute the asymptotics of the three auxiliary functions of Conjecture 6.1. Using eq. (D.42) and the specific form of the channel, we reach:

$$I_0(q_x) + I_{\text{int}}(q_x, q_z) \sim -\frac{\beta(\alpha - 1 + \nu(\{0\}))}{2} \ln \Delta,$$
$$I_{\text{out}}(q_z) \sim \frac{(\beta - 1)}{2} \ln \Delta.$$

Therefore when considering the total free entropy we have

$$I_0(q_x) + I_{\text{int}}(q_x, q_z) + \alpha I_{\text{out}}(q_z) \sim \frac{\beta(1-\nu(\{0\})) - \alpha}{2} \ln \Delta.$$

This implies that the full recovery point has a free entropy of $-\infty$ for $\alpha < \alpha_{\text{FR,IT}} \equiv \beta(1-\nu(\{0\}))$, and $+\infty$ for $\alpha > \alpha_{\text{FR,IT}}$. Thus this point is always the global maximum of the free entropy for $\alpha > \alpha_{\text{FR,IT}}$, while it is never the case for $\alpha < \alpha_{\text{FR,IT}}$, which ends our argument.

D.6 Details of the spectral methods analysis for phase retrieval

D.6.1 The expansion of $F(x, y)$ around $y = 0$

We describe here the behavior of $F(x, y)$ defined in eq. (6.25) as $x > 0$ and $y \downarrow 0$. Let us write the equations satisfied by ζ_x, ζ_y:

$$\Big\langle \frac{\zeta_y}{\zeta_x\zeta_y + \lambda} \Big\rangle_\nu = x, \qquad \frac{\alpha - 1}{\zeta_y} + \Big\langle \frac{\zeta_x}{\zeta_x\zeta_y + \lambda} \Big\rangle_\nu = \alpha y. \tag{D.43}$$

As $y \downarrow 0$, this implies necessarily that $\zeta_y \to +\infty$, and one finds easily that $\zeta_y \sim 1/y$, $\zeta_x \sim 1/x$. We now turn to the next order variations, that we write as:

$$\zeta_y = y^{-1} + c_1 + \mathcal{O}(y), \qquad \zeta_x = \frac{1}{x} + c_2 y + \mathcal{O}(y^2).$$

We use eq. (D.43) to compute $c_1 = -x\langle\lambda\rangle_\nu/\alpha$ and $c_2 = -\langle\lambda\rangle_\nu$. We can then develop:

$$\frac{1}{2}\langle \ln(\zeta_x\zeta_y + \lambda)\rangle_\nu = -\frac{1}{2}\ln y - \frac{1}{2}\ln x - \frac{x}{2\alpha}\langle\lambda\rangle_\nu y + \mathcal{O}(y^2).$$

Developing the other terms involved in $F(x, y)$ is straightforward and yields:

$$F(x, y) = -\frac{xy}{2}\langle\lambda\rangle_\nu + \mathcal{O}(y^2). \tag{D.44}$$

One can push this analysis to the next order, and finds in the exact same way, from eq. (D.43):

$$\zeta_y = \frac{1}{y} - \frac{x\langle\lambda\rangle_\nu}{\alpha} + \frac{x^2}{\alpha^2}\Big[\alpha\langle\lambda^2\rangle_\nu - (1+\alpha)\langle\lambda\rangle_\nu^2\Big]y + \mathcal{O}(y^2),$$
$$\zeta_x = \frac{1}{x} - \langle\lambda\rangle_\nu y + \frac{x}{\alpha}\Big[\alpha\langle\lambda^2\rangle_\nu - (1+\alpha)\langle\lambda\rangle_\nu^2\Big]y^2 + \mathcal{O}(y^3).$$

This yields for $F(x, y)$:

$$F(x, y) = -\frac{\langle\lambda\rangle_\nu}{2}xy + \frac{x^2}{4\alpha}[\alpha\langle\lambda^2\rangle_\nu - (1+\alpha)\langle\lambda\rangle_\nu^2]y^2 + \mathcal{O}(y^3),$$

which concludes our analysis.

D.6.2 Proof of Proposition 6.6

Let us recall the two spectral methods $\mathbf{M}^{(\mathrm{TAP})}$, $\mathbf{M}^{(\mathrm{LAMP})}$. Without loss of generality, we assume $\langle\lambda\rangle_\nu = \alpha$. Recall that we defined $z_\mu \equiv \partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho)$. We let $\mathbf{Z} \equiv \mathrm{Diag}(z_\mu)$. We can thus write:

$$\mathbf{M}^{(\mathrm{LAMP})} = \rho\Big(\frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n} - \mathrm{I}_m\Big)\mathbf{Z} \quad \text{and} \quad \mathbf{M}^{(\mathrm{TAP})} = -\frac{1}{\rho}\mathrm{I}_n + \frac{1}{n}\mathbf{\Phi}^\dagger\frac{\mathbf{Z}}{\mathrm{I}_m + \rho\mathbf{Z}}\mathbf{\Phi}.$$

We start by the first claim. By definition of $(\lambda_{\mathrm{LAMP}}, \mathbf{v})$, we have

$$\rho\frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n}\mathbf{Z}\mathbf{v} = (\rho\mathbf{Z} + \lambda_{\mathrm{LAMP}})\mathbf{v}. \tag{D.45}$$

Since we assumed $\lambda_{\text{LAMP}} + \rho z_\mu \neq 0$ for all μ, this implies that $\mathbf{\Phi}^\dagger \mathbf{Z} \mathbf{v} \neq 0$, and we thus let

$$\hat{\mathbf{x}} \equiv \frac{\mathbf{\Phi}^\dagger \mathbf{Z} \mathbf{v}}{\|\mathbf{\Phi}^\dagger \mathbf{Z} \mathbf{v}\|} \sqrt{n}.$$

Multiplying eq. (D.45) by $\mathbf{\Phi}^\dagger \mathbf{Z} (\lambda_{\text{LAMP}} + \rho \mathbf{Z})^{-1}$ on both sides, we directly reach the sought result:

$$\Big\{ \frac{1}{n} \mathbf{\Phi}^\dagger \frac{\rho \mathbf{Z}}{\lambda_{\text{LAMP}} + \rho \mathbf{Z}} \mathbf{\Phi} \Big\} \hat{\mathbf{x}} = \hat{\mathbf{x}}.$$

We move on to the second claim. Let $\mathbf{x} \in \mathbb{K}^n$ be an eigenvector of $\mathbf{M}^{(\text{TAP})}$ with norm $\|\mathbf{x}\|^2 = n$, with associated eigenvalue λ_{TAP}. We let:

$$\mathbf{u} \equiv \frac{\mathrm{I}_m}{\mathrm{I}_m + \rho \mathbf{Z}} \frac{\mathbf{\Phi}}{\sqrt{n}} \mathbf{x}.$$

And we can then easily compute:

$$\begin{aligned} \mathbf{M}^{(\text{LAMP})} \mathbf{u} &= \rho \Big(\frac{\mathbf{\Phi} \mathbf{\Phi}^\dagger}{n} - \mathrm{I}_m \Big) \frac{\mathbf{Z}}{\mathrm{I}_m + \rho \mathbf{Z}} \frac{\mathbf{\Phi}}{\sqrt{n}} \mathbf{x} = \frac{\rho \mathbf{\Phi}}{\sqrt{n}} \Big[\mathbf{M}^{(\text{TAP})} + \frac{1}{\rho} \mathrm{I}_n \Big] \mathbf{x} - \rho \mathbf{Z} \mathbf{u}, \\ &= \rho \lambda_{\text{TAP}} \frac{\mathbf{\Phi}}{\sqrt{n}} \mathbf{x} + \frac{\mathbf{\Phi}}{\sqrt{n}} \mathbf{x} - \rho \mathbf{Z} \mathbf{u} = \mathbf{u} + \rho \lambda_{\text{TAP}} (\mathrm{I}_m + \rho \mathbf{Z}) \mathbf{u}. \end{aligned} \tag{D.46}$$

At $\alpha = \alpha_{\text{WR,Algo}}$, the largest eigenvalue of $\mathbf{M}^{(\text{TAP})}$ concentrates on 0, which corresponds to the onset of marginal instability of the trivial local maximum. As one can see from eq. (D.46), this implies that $\mathbf{M}^{(\text{LAMP})}$ also possesses an eigenvalue equal to 1 at $\alpha = \alpha_{\text{WR,Algo}}$, indicating marginal instability as well. To put it shortly, *the two spectral methods have the same weak recovery threshold*. Moreover, eq. (D.46) implies that for any $\alpha \geq \alpha_{\text{WR,Algo}}$, if $\mathbf{M}^{(\text{TAP})}$ has en eigenvalue that concentrates on 0 as $n \to \infty$, then $\mathbf{M}^{(\text{LAMP})}$ has a corresponding eigenvalue concentrating on 1, and *with the same performance*. Indeed, as described in eq. (6.24), the estimator associated to $\mathbf{M}^{(\text{LAMP})}$ will be given by:

$$\hat{\mathbf{x}}_{\text{LAMP}} \propto \frac{\mathbf{\Phi}^\dagger}{\sqrt{n}} \mathbf{Z} \mathbf{u} = \frac{\mathbf{\Phi}^\dagger}{\sqrt{n}} \frac{\mathbf{Z}}{\mathrm{I}_m + \rho \mathbf{Z}} \frac{\mathbf{\Phi}}{\sqrt{n}} \hat{\mathbf{x}}_{\text{TAP}},$$

in which $\hat{\mathbf{x}}_{\text{TAP}}$ is an eigenvector of $\mathbf{M}^{(\text{TAP})}$ with eigenvalue 0. Therefore, we reach that $\hat{\mathbf{x}}_{\text{LAMP}} \propto \hat{\mathbf{x}}_{\text{TAP}}$, and these two vectors are thus equal as they are both normalized.

Appendix E

Details of the topological approach

E.1 The quenched complexity calculation

E.1.1 The phase volume factor

Introducing the Fourier transform of the deltas, we reach at leading exponential order in n:

$$\frac{1}{n}\ln\prod_{a=1}^{p}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{x}^a\prod_{a\leq b}\delta\big(nq_{ab}-n\mathbf{x}^a\cdot\mathbf{x}^b\big)\simeq\frac{p}{2}\ln\frac{2\pi}{n}+\frac{1}{2}\sup_{\{\hat{q}_{ab}\}}\Big[\sum_{a,b}q_{ab}\hat{q}_{ab}-\ln\det\hat{\mathbf{q}}\Big].$$

The replica symmetric assumption can also be made on the variables $\hat{\mathbf{q}}$ that achieve this supremum : $\hat{q}_{aa}=\hat{q}_0$ and $\hat{q}_{ab}=-\hat{q}$ for $a\neq b$. This leads to $\det\hat{\mathbf{q}}=(\hat{q}_0+\hat{q})^{p-1}(\hat{q}_0-(p-1)\hat{q})$, and after taking the $p\downarrow 0$ limit, we reach:

$$\frac{1}{np}\ln\prod_{a=1}^{p}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{x}^a\prod_{a\leq b}\delta\big(nq_{ab}-n\mathbf{x}^a\cdot\mathbf{x}^b\big)\simeq\frac{1}{2}\ln\frac{2\pi}{n}+\frac{1}{2}\sup_{\hat{q}_0,\hat{q}}\Big[\hat{q}_0+q\hat{q}-\ln(\hat{q}_0+\hat{q})+\frac{\hat{q}}{\hat{q}_0+\hat{q}}\Big].$$

The diverging term $-(\ln n)/2$ will be canceled out by the joint density of the gradients as we will see later. The solution of the supremum is easy to carry out, and we finally reach eq. (7.30).

E.1.2 The joint density of the gradients

We denote $S=\mathrm{Span}\,(\{\mathbf{x}^a\}_{a=1}^p)\subset\mathbb{R}^n$. Following [RBABC19], for every $1\leq a\leq p$ we can construct an orthonormal basis of S, denoted $(\mathbf{e}_b^a)_{1\leq b\leq p}$ for which $\mathbf{x}^a$ is the first vector, that is $\mathbf{e}_a^a=\mathbf{x}^a$. This basis is convenient, since $\{\mathbf{x}^a\}^\perp\cap S=\mathrm{Span}\,(\{\mathbf{e}_b^a\}_{b(\neq a)})$. We can also chose an arbitrary orthonormal basis $(\mathbf{e}_{p+1},\cdots,\mathbf{e}_n)$ of $S^\perp$. With this choice of basis, we can see that the gradient $\mathrm{grad}\,L(\mathbf{x}^a)$ is identified with the vector in $\mathbb{R}^{n-1}$ with components:

$$\mathrm{grad}\,L(\mathbf{x}^a)=(\{\nabla L(\mathbf{x}^a)\cdot\mathbf{e}_i^a\}_{i=1}^{a-1},\{\nabla L(\mathbf{x}^a)\cdot\mathbf{e}_i^a\}_{i=a+1}^{p},\{\nabla L(\mathbf{x}^a)\cdot\mathbf{e}_i\}_{i=p+1}^{n}).\tag{E.1}$$

Recall that $\nabla L(\mathbf{x}^a)=(1/m)\sum_\mu\boldsymbol{\xi}_\mu\,\phi'(y_\mu^a)$. Let us make a few remarks:

- For every a, the Gram matrix of the basis $(\mathbf{e}_b^a)_{b=1}^p$ is only a function of the values of the overlaps $\{q_{ab}\}$.
- We consider the joint density of the gradients *conditioned* by the value of $\{\mathbf{y}^a\}$. In particular, this means that for every $a\neq b$, $\nabla L(\mathbf{x}^a)\cdot\mathbf{e}_b^a$ is fixed by the values of $\{\mathbf{y}^c\}_{c=1}^p$ and the overlaps q_{ab}. In particular, the first $(p-1)$ components of eq. (E.1) are deterministic, thus their density will yield delta functions that are constraints on $\{\mathbf{y}^a\}$ and $\{q_{ab}\}$.
- The last $n-p$ components of eq. (E.1) are (at fixed $\{\mathbf{y}^a\}$) zero mean Gaussian random variables with covariance given by $\mathbb{E}[\mathrm{grad}L(\mathbf{x}^a)_i\,\mathrm{grad}L(\mathbf{x}^b)_j]=(\delta_{ij}/m^2)\sum_\mu\phi'(y_\mu^a)\,\phi'(y_\mu^b)$. Their joint

density taken at 0 is thus at leading exponential order in n:

$$\exp\left\{\frac{np}{2}\ln\frac{m}{2\pi}-\frac{n}{2}\ln\det\left[\left(\frac{1}{m}\sum_{\mu=1}^{m}\phi'(y_\mu^a)\phi'(y_\mu^b)\right)_{1\le a,b\le p}\right]\right\}. \tag{E.2}$$

Given these remarks and eq. (E.2), in order to complete the calculation of the joint gradient density we need to compute the quantities $(\nabla L(\mathbf{x}^a)\cdot\mathbf{e}_b^a)$ for every $a\neq b$ as a function of $\{y_\mu^a\}$ and $\{q_{ab}\}$. In order to simplify the calculation, we will already make use of the replica-symmetric assumption on q, that is we assume $q_{aa}=1$ and $q_{ab}=q$ for $a\neq b$. Let us now describe a possible construction for the basis $(\mathbf{e}_b^a)_{b=1}^p$. We introduce three auxiliary quantities that are functions of q and p:

$$\begin{cases} f_p^0(q) &\equiv \dfrac{1}{p-1}\Big[\dfrac{p-2}{\sqrt{1-q}}+\dfrac{1}{\sqrt{1+(p-2)q-(p-1)q^2}}\Big],\\ f_p(q) &\equiv \dfrac{1}{p-1}\Big[-\dfrac{1}{\sqrt{1-q}}+\dfrac{1}{\sqrt{1+(p-2)q-(p-1)q^2}}\Big],\\ z_p(q) &\equiv -\dfrac{q}{\sqrt{1+(p-2)q-(p-1)q^2}}. \end{cases} \tag{E.3}$$

Using these definitions, we can consider:

$$\begin{cases} \mathbf{e}_a^a &\equiv \mathbf{x}^a,\\ \mathbf{e}_b^a &\equiv z_p(q)\mathbf{x}^a+f_p^0(q)\mathbf{x}^b+f_p(q)\sum_{c(\neq a,b)}\mathbf{x}^c, \qquad (b\neq a). \end{cases} \tag{E.4}$$

It is straightforward to check from eq. (E.4) that we have for all a,b,c that $\mathbf{e}_b^a\cdot\mathbf{e}_c^a=\delta_{bc}$. We can now see that the delta term of the joint density of the gradients taken at 0 is:

$$\prod_{a\neq b}\delta[\nabla L_1(\mathbf{x}^a)\cdot\mathbf{e}_b^a]=\prod_{a\neq b}\delta\left[\frac{1}{m}\sum_{\mu=1}^{m}\phi'(y_\mu^a)\Big(z_p(q)y_\mu^a+f_p^0(q)y_\mu^b+f_p(q)\sum_{c(\neq a,b)}y_\mu^c\Big)\right]. \tag{E.5}$$

The product of eq. (E.2) and eq. (E.5) gives eq. (7.31).

E.1.3 Decoupling replicas, and the $p\downarrow 0$ limit

Replica symmetry and decoupling

In order to apply the replica method, we need to be able to take the $p\downarrow 0$ limit, by analytically extending eq. (7.35) to all $p>0$. The main idea is that we expect replica symmetry to influence the measure ν that solves the supremum in eq. (7.35). More precisely, we expect that for all permutation π of $\{1,\cdots,p\}$, we have $\nu(\mathrm{d}\lambda^1,\cdots,\mathrm{d}\lambda^p)=\nu(\mathrm{d}\lambda^{\pi(1)},\cdots,\mathrm{d}\lambda^{\pi(p)})$. Let us see how this hypothesis simplifies the calculation. We separate in eq. (7.35) the *marginals* of ν, in the following way:

$$\sup_{\nu\in\mathcal{M}(p,q)} \;\to\; \sup_{\{\mu_a\}_{a=1}^p\in\mathcal{M}_1^+(\mathbb{R})}\ \sup_{\substack{\nu\in\mathcal{M}(p,q)\\ \text{s.t. }\{\nu^a=\mu_a\}}} \tag{E.6}$$

In this last expression, the replica symmetric assumption leads us in particular to assume that $\mu_a=\mu$ for all a. In order to make the remaining calculation tractable we will also need to fix some linear statistics of ν via Lagrange multipliers:

- For every $a \leq b$, we fix the linear statistics $\int \nu(\mathrm{d}\lambda)\, \phi'(\lambda^a)\phi'(\lambda^b) = A_{ab}$, with Lagrange multipliers $\hat{A}_{ab}$. Note that by replica symmetry, we can assume that $A_{ab} = a$ for $a \neq b$ and $A_{aa} = A$ (and similarly for the Lagrange multipliers).
- For all a, b we fix the linear statistics $\int \nu(\mathrm{d}\lambda)\, \phi'(\lambda^a)\lambda^b = B_{ab}$, with Lagrange multipliers $\hat{B}_{ab}$. By replica symmetry, we assume that $B_{aa} = B$ and $B_{ab} = b$ (and similarly for $\hat{B}_{ab}$).

Combining these remarks, we reach that the ν-dependent term of eq. (7.35) is equal to:

$$\sup_{\mu\in\mathcal{M}_\phi(B)} \sup_{\substack{A,a\\B,b}} \operatorname*{extr}_{\substack{\hat{A},\hat{a}\\\hat{B},\hat{b}}} \sup_{\substack{\nu\in\mathcal{M}_1^+(\mathbb{R}^n)\\ \text{s.t. } \{\nu^a=\mu\}}} \Bigg\{ p\kappa_{\alpha,\phi}[\mu, t_\phi(\mu)] - \frac{1}{2}\ln\det[\{A_{ab}\}] - \sum_{a,b}\Big[\frac{1}{2}A_{ab}\hat{A}_{ab} + B_{ab}\hat{B}_{ab}\Big]$$
$$+ \sum_{a,b}\Big[\frac{1}{2}\hat{A}_{ab}\int \nu(\mathrm{d}\lambda)\phi'(\lambda^a)\phi'(\lambda^b) + \hat{B}_{ab}\int \nu(\mathrm{d}\lambda)\phi'(\lambda^a)\lambda^b\Big] - \alpha D_{\mathrm{KL}}(\nu|\mu_{\mathrm{G},q})\Bigg\}. \tag{E.7}$$

Note that here we did not always explicit the replica-symmetry assumption on all the variables to obtain more compact expressions. The supremum over B, b is moreover constrained by the following condition of eq. (7.34b): $\forall a \neq b,\ z_p(q)B_{aa} + f_p^0(q)B_{ab} + f_p(q)\sum_{c(\neq a,b)} B_{ac} = 0$. Under the replica symmetric assumption, this becomes:

$$z_p(q)B + f_p^0(q)b + f_p(q)(p-2)b = 0. \tag{E.8}$$

Again, we introduce Lagrange multipliers C_{ab} to fix these conditions, that reduce to $C_{ab} = C$ because of replica symmetry. Finally, in order to fix the marginal distributions of ν, we will have to introduce "functional" Lagrange multipliers $g^a(\lambda^a)$. Again, by replica symmetry, we expect all of them to be equal to $g(\lambda^a)$. In the end, we obtain from eq. (E.7):

$$\sup_{\substack{\mu\in\mathcal{M}_\phi(B)\\ \nu\in\mathcal{M}(\mathbb{R}^n)}} \sup_{\substack{A,a\\B,b}} \operatorname*{extr}_{\substack{C,\hat{A},\hat{a}\\\hat{B},\hat{b},\{g(\lambda)\}}} \Bigg\{ p\kappa_{\alpha,\phi}[\mu, t_\phi(\mu)] - \frac{1}{2}\ln\det[\{A_{ab}\}] - \sum_{a,b}\Big[\frac{1}{2}\hat{A}_{ab}A_{ab} + \hat{B}_{ab}B_{ab}\Big] \tag{E.9}$$
$$- p\int \mu(\mathrm{d}\lambda)g(\lambda) + p(p-1)C\big[Bz_p(q) + b\{f_p^0(q) + (p-2)f_p(q)\}\big] - \alpha D_{\mathrm{KL}}(\nu|\mu_{\mathrm{G},q})$$
$$+ \sum_{a,b}\Big[\frac{\hat{A}_{ab}}{2}\int \nu(\mathrm{d}\lambda)\phi'(\lambda^a)\phi'(\lambda^b) + \hat{B}_{ab}\int \nu(\mathrm{d}\lambda)\phi'(\lambda^a)\lambda^b\Big] + \sum_a \int \nu(\mathrm{d}\lambda)g(\lambda^a)\Bigg\}.$$

We can now solve exactly the supremum over ν. By a classical Gibbs measure calculation that we already detailed in Section 7.3 we obtain (recall that $\mathbf{Q} \in \mathcal{S}_p$ is the overlap matrix):

$$\sup_{\nu\in\mathcal{M}(\mathbb{R}^n)} \Bigg\{ \int \nu(\mathrm{d}\lambda)\Big[\sum_{a,b}\Big(\frac{\hat{A}_{ab}}{2}\phi'(\lambda^a)\phi'(\lambda^b) + \hat{B}_{ab}\phi'(\lambda^a)\lambda^b\Big) + \sum_a g(\lambda^a)\Big] - \alpha D_{\mathrm{KL}}(\nu|\mu_{\mathrm{G},q})\Bigg\}$$
$$= \alpha \ln \int_{\mathbb{R}^p} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}^p\sqrt{\det\mathbf{Q}}} e^{\sum_{a,b}\left(-\frac{1}{2}(\mathbf{Q}^{-1})_{ab}\lambda^a\lambda^b + \frac{\hat{A}_{ab}}{2\alpha}\phi'(\lambda^a)\phi'(\lambda^b) + \frac{\hat{B}_{ab}}{\alpha}\phi'(\lambda^a)\lambda^b\right) + \sum_a \frac{g(\lambda^a)}{\alpha}}. \tag{E.10}$$

To completely decouple the replicas, we will make use of two classical identities, for any x, y:

$$e^{\frac{x^2}{2}} = \int \mathcal{D}\xi\, e^{\xi x}\,, \qquad e^{xy} = \int \mathcal{D}\xi\mathcal{D}\xi'\, e^{\frac{x}{\sqrt{2}}(\xi+i\xi') + \frac{y}{\sqrt{2}}(\xi-i\xi')}.$$

Thanks to replica symmetry, we can compute $\mathbf{Q}^{-1}$ and $\det \mathbf{Q}$ as:

$$\begin{cases} \det \mathbf{Q} & = (1-q)^{p-1}[1+(p-1)q], \\ \mathbf{Q}^{-1}_{ab} & = \dfrac{1+(p-1)q}{1+(p-2)q-(p-1)q^2}\delta_{ab} - \dfrac{q}{(1-q)(1+(p-1)q)}. \end{cases}$$

We define

$$\begin{cases} d_{0,p}(q) & \equiv \dfrac{1+(p-1)q}{1+(p-2)q-(p-1)q^2}, \\ d_p(q) & \equiv \dfrac{q}{(1-q)(1+(p-1)q)}. \end{cases}$$

Using all the above, we can now simplify eq. (E.10):

$$\begin{aligned} &\alpha \ln \int_{\mathbb{R}^p} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}^p \sqrt{\det \mathbf{Q}}} e^{\sum_{a,b}\left(-\frac{1}{2}(\mathbf{Q}^{-1})_{ab}\lambda^a\lambda^b + \frac{\hat{A}_{ab}}{2\alpha}\phi'(\lambda^a)\phi'(\lambda^b) + \frac{\hat{B}_{ab}}{\alpha}\phi'(\lambda^a)\lambda^b\right) + \sum_a \frac{g(\lambda^a)}{\alpha}} \\ &= -\frac{\alpha p}{2}\ln 2\pi - \frac{\alpha(p-1)}{2}\ln(1-q) - \frac{\alpha}{2}\ln[1+(p-1)q] + \alpha \ln \int_{\mathbb{R}^4} \mathcal{D}\boldsymbol{\xi} I_p(\boldsymbol{\xi})^p, \end{aligned} \tag{E.11}$$

in which we defined $\boldsymbol{\xi} \equiv (\xi_q, \xi_a, \xi_b, \xi'_b)$ and

$$I_p(\boldsymbol{\xi}) \equiv \int \mathrm{d}\lambda\, e^{\frac{g(\lambda)}{\alpha} - \frac{d_{0,p}(q)\lambda^2}{2} + \frac{\hat{A}-\hat{a}}{2\alpha}\phi'(\lambda)^2 + \frac{\hat{B}-\hat{b}}{\alpha}\phi'(\lambda)\lambda + \sqrt{d_p(q)}\xi_q\lambda + \sqrt{\frac{\hat{a}}{\alpha}}\xi_a\phi'(\lambda) + \sqrt{\frac{\hat{b}}{2\alpha}}[\phi'(\lambda)(\xi_b+i\xi'_b)+\lambda(\xi_b-i\xi'_b)]}.$$

Although the involved expressions are very cumbersome, we have successfully decoupled the replicas.

The $p \downarrow 0$ limit, and the final result

We begin by a remark on eq. (E.11). Note that $\lim_{p\downarrow 0}(1/p)\ln \int \mathcal{D}\xi\, I_p(\boldsymbol{\xi})^p = \int \mathcal{D}\boldsymbol{\xi}\, \ln I_0(\boldsymbol{\xi})$. Thus, after multiplication by $(1/p)$, the $p \downarrow 0$ limit of eq. (E.11) will yield:

$$-\frac{\alpha}{2}\ln 2\pi - \frac{\alpha}{2}\ln(1-q) - \frac{\alpha q}{2(1-q)} + \alpha \int \mathcal{D}\boldsymbol{\xi} \ln I(\boldsymbol{\xi}), \tag{E.12}$$

in which $I(\boldsymbol{\xi})$ is defined in Result 7.1. We can wrap up the calculation. We make two remarks. First the condition of eq. (E.8) reduces, in the $p \downarrow 0$ limit, to $b = qB$, so that we will be able to simplify the terms involving the Lagrange multiplier C. Secondly, the variable B is equal to $t_\phi(\mu)$, defined in Theorem 7.5. We combine now eqs. (7.34a),(7.35),(E.9) and (E.12) with the two remarks above. Changing notations from μ to ν and B to C, we obtain finally the conclusion of Result 7.1.

E.2 Details of proof for the annealed complexity

E.2.1 Proof of Lemma 7.7

We will apply the Kac-Rice machinery in the form of the remark made in Paragraph 6.1.4 of [AW09b]. We recall it as a theorem:

Theorem E.1 (*Azais-Wschebor*)

Let $k, d \in \mathbb{N}^\star$. Let $Z : U \to \mathbb{R}^d$ be a random field, in which U is an open subset of $\mathbb{R}^d$. Assume that for every $t \in U$, we can write $Z(t) = H[Y(t)]$, such that:

(i) $\{Y(t), t \in U\}$ is a Gaussian random field with values in $\mathbb{R}^k$, $\mathcal{C}^1$ paths, and such that for every $t \in U$, the distribution of $Y(t)$ is non-degenerate.

(ii) $H : \mathbb{R}^k \to \mathbb{R}^d$ is a $\mathcal{C}^1$ function.

(iii) For all $t \in U$, $Z(t)$ has a density $\varphi_{Z(t)}(x)$ which is a continuous function of $(t, x) \in U \times \mathbb{R}^d$.

(iv) $\mathbb{P}[\exists\, t \in U \text{ s.t. } Z(t) = 0 \text{ and } \det \nabla Z(t) = 0] = 0$.

For every compact set $B \subseteq U$, we let $N(Z, B)$ the (finite) number of zeros of Z in B. Then:

$$\mathbb{E}[N(Z, B)] = \int_B \mathbb{E}[|\det \nabla Z(t)| | Z(t) = 0] \varphi_{Z(t)}(0) \mathrm{d}t.$$

We wish to apply this theorem to $Z(\mathbf{x}) = \operatorname{grad} L_1(\mathbf{x})$. Verifying its hypotheses will end the proof of Lemma 7.7. We denote $\boldsymbol{\xi} \in \mathbb{R}^{n \times m}$ the matrix $\{\xi_{i\mu}\} = \{(\boldsymbol{\xi}_\mu)_i\}$, ∇L_1 the Euclidean gradient of L_1, and $P_{\mathbf{x}}^\perp$ the orthogonal projection on $T_{\mathbf{x}}\mathbb{S}^{n-1}$. Since $\operatorname{grad} L_1(\mathbf{x}) = P_{\mathbf{x}}^\perp \nabla L_1(\mathbf{x})$ we have:

$$\operatorname{grad} L_1(\mathbf{x}) = \frac{1}{m} \sum_{\mu=1}^m (P_{\mathbf{x}}^\perp \boldsymbol{\xi}_\mu) \phi'(\boldsymbol{\xi}_\mu \cdot \mathbf{x}). \tag{E.13}$$

We will apply Theorem E.1 with $d = n - 1$ and $k = m \times n$. The Gaussian random field $Y(\mathbf{x}) \in \mathbb{R}^{n \times m}$ is defined as $Y(\mathbf{x}) \equiv \begin{pmatrix} P_{\mathbf{x}}^\perp \boldsymbol{\xi}_1 & \cdots & P_{\mathbf{x}}^\perp \boldsymbol{\xi}_m \\ \boldsymbol{\xi}_1 \cdot \mathbf{x} & \cdots & \boldsymbol{\xi}_m \cdot \mathbf{x} \end{pmatrix}$. Since $Y(\mathbf{x})$ is just $\boldsymbol{\xi}$ written in an orthonormal basis of $\mathbb{R}^n$ whose last vector is $\mathbf{x}$, its distribution is non-degenerate for every $\mathbf{x}$. $H : \mathbb{R}^{n \times m} \to \mathbb{R}^{n-1}$ is defined as:

$$\forall 1 \leq i < n, \quad H(Y)_i \equiv \frac{1}{m} \sum_{\mu=1}^m Y_{i,\mu} \phi'(Y_{n,\mu}), \quad (Y \in \mathbb{R}^{n \times m}).$$

Since ϕ is $\mathcal{C}^2$, H is $\mathcal{C}^1$. This verifies (i) and (ii). We turn our attention to verifying (iii). One can write the distribution of the gradient of eq. (E.13) as $\operatorname{grad} L_1(\mathbf{x}) \overset{\mathrm{d}}{=} (1/m) \sum_{\mu=1}^m \phi'(y_\mu) \mathbf{z}_\mu$, in which $y_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, $\mathbf{z}_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_{n-1})$, and all $\{y_\mu, \mathbf{z}_\nu\}$ are independent. Since the distribution of $\operatorname{grad} L_1(\mathbf{x})$ does not depend on $\mathbf{x}$, it is enough to check that its density exists and is a continuous function. To do so, we will show that its characteristic function $\hat{\varphi}_{\operatorname{grad} L_1(\mathbf{x})} \in L^1(\mathbb{R}^{n-1})$. We denote $\hat{\varphi}_a$ the characteristic function of the random variable $a \equiv \phi'(y)$, and one obtains:

$$\begin{aligned} \|\hat{\varphi}_{\operatorname{grad} L_1(\mathbf{x})}\|_1 &= \int_{\mathbb{R}^{n-1}} \mathrm{d}\mathbf{t} \left| \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(0, I_{n-1})} \hat{\varphi}_a\Big(\frac{\mathbf{t} \cdot \mathbf{z}}{m}\Big) \right|^m = \int_{\mathbb{R}^{n-1}} \mathrm{d}\mathbf{t} \left| \mathbb{E}_{z \sim \mathcal{N}(0,1)} \hat{\varphi}_a\Big(\frac{\|\mathbf{t}\| z}{m}\Big) \right|^m, \\ &= \frac{2\pi^{\frac{n-1}{2}} m^{n-1}}{\Gamma\big(\frac{n-1}{2}\big)} \int_0^\infty \mathrm{d}q\, q^{n-2} |\mathbb{E}_z \hat{\varphi}_a(qz)|^m. \end{aligned}$$

Since $\alpha > 1$, if $q \mathbb{E}_z \hat{\varphi}_a(qz) = \mathcal{O}_{q \to \infty}(1)$ we can conclude that $\|\hat{\varphi}_{\operatorname{grad} L_1(\mathbf{x})}\|_1 < \infty$. And:

$$q\, \mathbb{E}_z \hat{\varphi}_a(qz) = \int_{\mathbb{R}} \frac{\mathrm{d}z}{\sqrt{2\pi}} e^{-\frac{z^2}{2q^2}} \hat{\varphi}_a(z) = \int_{\mathbb{R}} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathbb{E}\Big[e^{-\frac{z^2}{2q^2}} e^{iaz}\Big] = \frac{1}{q} \mathbb{E}\Big[e^{-\frac{q^2 a^2}{2}}\Big],$$

by Fubini's theorem. Therefore $q\,\mathbb{E}_z\,\hat{\varphi}_a(qz) \underset{q\to\infty}{\to} \varphi_a(0)$ by continuity of φ_a around $a = 0$ (Definition 7.1), so $\|\hat{\varphi}_{\mathrm{grad}L_1(\mathbf{x})}\|_1 < \infty$. Thus $\mathrm{grad}L_1(\mathbf{x})$ admits the following probability density:

$$\varphi_{\mathrm{grad}L_1(\mathbf{x})}(\mathbf{u}) = \frac{1}{(2\pi)^{n-1}} \int_{\mathbb{R}^{n-1}} \mathrm{d}\mathbf{t}\, e^{i\mathbf{u}\cdot\mathbf{t}} \Big[\mathbb{E}_z\Big\{\hat{\varphi}_a\Big(\frac{\|\mathbf{t}\|z}{m}\Big)\Big\}\Big]^m, \tag{E.14}$$

which is a continuous function of $\mathbf{u}$, since $\hat{\varphi}_{\mathrm{grad}L_1(\mathbf{x})} \in L^1(\mathbb{R}^{n-1})$. This shows (iii). In order to show (iv), we will use Proposition 6.5 of [AW09b], that we recall here:

Lemma E.2 (*Azais-Wschebor*)

Let $d \in \mathbb{N}^*$, and U a compact subset of $\mathbb{R}^d$. Consider $Z : U \to \mathbb{R}^d$ a random field, such that (a): The paths of Z are of class $\mathcal{C}^2$, and (b): There exists $C > 0$ such that for all $t \in U$ and all u in a neighborhood of 0, the density $\varphi_{Z(t)}$ of Z verifies $\varphi_{Z(t)}(u) \le C$. Then $\mathbb{P}[\exists t \in U \text{ s.t. } Z(t) = 0 \text{ and } \det Z'(t) = 0] = 0$.

Since ϕ is assumed to be of class $\mathcal{C}^3$, hypothesis (a) is verified for $Z = \mathrm{grad}\, L_1$. Notice then that we can fix $C > 0$ such that $|\mathbb{E}_{z\sim\mathcal{N}(0,1)}\,\hat{\varphi}_a(qz)| \le C/(1+q)$ for all $q \ge 0$. Starting from eq. (E.14):

$$|\varphi_{\mathrm{grad}L_1(\mathbf{x})}(\mathbf{u})| \le C_n \int_0^\infty \mathrm{d}q\, \frac{q^{n-2}}{(1+q)^m} \le D_n,$$

with C_c, D_n constants depending only on n, using that $m \ge n$ ($\alpha > 1$). This shows (b), so by Lemma E.2, hypothesis (iv) of Theorem E.1 follows. This ends the proof of Lemma 7.7.

E.2.2 Proof of Lemma 7.10

The proof is done in several parts, and is inspired by arguments of [Sil95, SB95, SC95, BS10].

Technicalities on the Hessian

We begin by a technical lemma on $\mathbf{\Lambda}(\mathbf{y})$, defined in eq. (7.12).

Lemma E.3 (*Low-rank perturbation*)

Since the distributions of $\mathbf{z}$ and $\mathbf{y}$ are independent, by rotation invariance we can assume that $\mathbf{\Lambda}(\mathbf{y})$ is a diagonal matrix with elements $\mathbf{\Lambda}_\mu(\mathbf{y})$. There exists a constant, denoted $||\mathbf{D}||_\infty$, such that for all n, y, $|D(y)| \le ||\mathbf{D}||_\infty$. Then we have:

(i) $\sup_{\mathbf{y}\in\mathbb{R}^m} \sup_{1\le\mu\le m} |\mathbf{\Lambda}_\mu(\mathbf{y})| \le 4||\mathbf{D}||_\infty$.

(ii) Let $\mathbf{Z} \in \mathbb{R}^{(n-1)\times m}$ be i.i.d. variables with zero mean and unit variance. We denote $\mu_{\mathbf{D}}^{(n)}$ and $\mu_{\mathbf{\Lambda}}^{(n)}$ the ESDs of $\mathbf{Z}\mathbf{D}(\mathbf{y})\mathbf{Z}^\intercal/n$ and $\mathbf{Z}\mathbf{\Lambda}(\mathbf{y})\mathbf{Z}^\intercal/n$ respectively. Then for all $\eta \in (0,1)$, $\{n^\eta \mathbb{E}_{\mathbf{z}}[\mu_D^{(n)} - \mu_{\mathbf{\Lambda}}^{(n)}]\} \underset{n\to\infty}{\to} 0$ weakly and uniformly in $\mathbf{y} \in \mathbb{R}^m$.

Proof of Lemma E.3 – Recall that $|D(y)| = (n/m)|\phi''(y)|$. Since $m/n \to \alpha > 1$ and ϕ'' is bounded, $|D(y)|$ is bounded (uniformly over $n, \mathbf{y}$) by a constant that we denote $||\mathbf{D}||_\infty$. Note that $\sup_{1\le\mu\le m} |\mathbf{\Lambda}_\mu(\mathbf{y})| = \sup_{||\mathbf{u}||=1} \mathbf{u}^\intercal\mathbf{\Lambda}(\mathbf{y})\mathbf{u}$. Using eq. (7.12) with $\mathbf{v}(\mathbf{y}) \equiv \phi'(\mathbf{y})/||\phi'(\mathbf{y})||$ we reach

$$\begin{aligned}\sup_{\|\mathbf{u}\|=1} \mathbf{u}^\intercal\mathbf{\Lambda}(\mathbf{y})\mathbf{u} &\le ||\mathbf{D}||_\infty + \sup_{\|\mathbf{u}\|=1} \Big[|\mathbf{v}^\intercal\mathbf{D}\mathbf{v}|(\mathbf{u}^\intercal\mathbf{v})^2 + 2(\mathbf{u}^\intercal\mathbf{v})|\mathbf{v}^\intercal\mathbf{D}\mathbf{u}|\Big],\\ &\le 2||\mathbf{D}||_\infty + 2 \sup_{||\mathbf{u}||=1} [(\mathbf{u}^\intercal\mathbf{v})|\mathbf{v}^\intercal\mathbf{D}\mathbf{u}|] \le 4\|\mathbf{D}\|_\infty,\end{aligned}$$

in which we used the uniform boundedness of $|D(y_\mu)|$, and the Cauchy-Schwarz inequality. This proves (i). We note that $\mathbf{z}\mathbf{\Lambda}(\mathbf{y})\mathbf{z}^\intercal/n$ and $\mathbf{z}\mathbf{D}(\mathbf{y})\mathbf{z}^\intercal/n$ differ by a rank-2 matrix. (ii) is thus an immediate application of the following result (Lemma 2.5 of [Bor19]):

Lemma E.4 (*Low-rank perturbation and empirical spectral distribution*)

Let $n \geq 1$, and $\mathbf{A}, \mathbf{B}$ two symmetric matrices of size n, such that the rank of $\mathbf{A} - \mathbf{B}$ is r. Denote $F_{\mathbf{A}}$ (resp. $F_{\mathbf{B}}$) the CDF of the empirical spectral distribution of $\mathbf{A}$ (resp. $\mathbf{B}$). Then

$$\sup_{t\in\mathbb{R}} |F_{\mathbf{A}}(t) - F_{\mathbf{B}}(t)| \leq \frac{r}{n} \ .$$

This ends the proof of Lemma E.3. □

Proof of Lemma E.4 – We note $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ the eigenvalues of $\mathbf{A}$ (and similarly for $\mathbf{B}$). Recall the weak Weyl's interlacing inequalities [Wey12]: for every $1 \leq i \leq n$, $\lambda_{i+r}(\mathbf{A}) \leq \lambda_i(\mathbf{B}) \leq \lambda_{i-r}(\mathbf{A})$ (we use the convention $\lambda_{1-i} = +\infty$ and $\lambda_{n+i} = -\infty$ for $i \geq 1$). Let $t \in \mathbb{R}$, and i, j be the smallest indices such that $\lambda_i(\mathbf{A}) \leq t$ and $\lambda_j(\mathbf{B}) < t$. By the interlacing inequalities, $|i - j| \leq r$. And $n|F_{\mathbf{A}}(t) - F_{\mathbf{B}}(t)| = |(n + 1 - i) - (n + 1 - j)| \leq r$. □

We have some control of the boundedness of the Hessian, summarized in two subsequent lemmas:

Lemma E.5 (*Moment bound*)

For all $\gamma > 0$, one has:

$$\limsup_{n\to\infty} \sup_{\mathbf{y}\in\mathbb{R}^m} \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}}\{|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|^\gamma\} < +\infty.$$

Proof of Lemma E.5 – We begin by bounding the extremal eigenvalues $\lambda_{\min}, \lambda_{\max}$ of $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$ (denoted $\mathbf{H}$ for lightness):

$$\begin{aligned}\lambda_{\max} &= \sup_{\|\mathbf{u}\|^2=1} [\mathbf{u}^\intercal \mathbf{H} \mathbf{u}] = -\frac{1}{m}\sum_{\mu=1}^m y_\mu \phi'(y_\mu) + \sup_{\|\mathbf{u}\|^2=1} \Big[\frac{1}{n}\sum_{\mu=1}^m \Lambda_\mu(\mathbf{y})(\mathbf{z}^\intercal\mathbf{u})_\mu^2\Big], \\ &\leq \|x\phi'(x)\|_\infty + 4\|\mathbf{D}\|_\infty \times \lambda_{\max}\Big[\frac{1}{n}\mathbf{z}\mathbf{z}^\intercal\Big].\end{aligned}$$

We used (i) of Lemma E.3. Note that this bound is independent of $\mathbf{y}$. In the same way we can bound $\lambda_{\min}$, and we reach:

$$\max(-\lambda_{\min}, \lambda_{\max}) \leq \|x\phi'(x)\|_\infty + 4\|\mathbf{D}\|_\infty \times \lambda_{\max}\Big[\frac{1}{n}\mathbf{z}\mathbf{z}^\intercal\Big],$$

Using this identity, we have the bound:

$$\frac{1}{n}\ln \mathbb{E}_{\mathbf{z}}\{|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|^\gamma\} \leq \frac{1}{n}\ln \mathbb{E}_{\mathbf{z}} \exp\Big\{n\gamma \ln\Big(\|x\phi'(x)\|_\infty + 4\|\mathbf{D}\|_\infty \lambda_{\max}(\mathbf{z}\mathbf{z}^\intercal/n)\Big)\Big\} \tag{E.15}$$

It is then a very classical result of random matrix theory that the largest eigenvalue of a Wishart matrix $\mathbf{z}\mathbf{z}^\intercal/n$ satisfies a large deviation principle in the scale n. This is stated e.g. in Theorem 2.4 of [BG20], or as a particular case of Chapter 8 of the present thesis (Result 8.1), which gives moreover the behavior of the rate function $I(x)$. We recall some of its properties:

- $I(x) = +\infty$ if $x < s_{\max} \equiv (1 + \alpha^{-1/2})^2$.
- $I(x) : [s_{\max}, +\infty) \to \mathbb{R}_+$ is continuous and increasing.

- $I(x) \sim_{x\to\infty} x/2$.

In particular, by Varadhan's lemma 1.10 it implies that for any $C, D, \gamma > 0$:

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}} \exp\left\{ n\gamma \ln\left[C + D\lambda_{\max}(\mathbf{z}\mathbf{z}^\intercal/n) \right] \right\} < +\infty.$$

Combining this inequality with eq. (E.15) ends the proof of Lemma E.5. □

Lemma E.6 (*Properties of $\mu_{\alpha,\phi}$*)

Denote $\rho_n(\mathbf{y})$ the spectral radius of $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$. There exists $C > 0$ such that:

(i) With probability 1, $\limsup_{n\to\infty} \sup_{\mathbf{y}\in\mathbb{R}^m} \rho_n(\mathbf{y}) < C$.

(ii) The support of $\mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]$ is included in $(-C, C)$ uniformly over $\mathbf{y}$ and n.

(iii) For all $\mathbf{y} \in \mathbb{R}^m$, $\mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]$ has a well-defined and continuous density outside $x = 0$.

Proof of Lemma E.6 – Points (ii) and (iii) are consequences of Theorem 1.1 of [SC95], while (i) follows from the boundedness of $\mathbf{\Lambda}(\mathbf{y})$ and the one of $x\phi'(x)$, as in Lemma E.5. □

The cut-off and the logarithmic potential

For any $\epsilon > 0$, define $\ln_\epsilon : x \in \mathbb{R}_+^\star \mapsto \ln(\max(x, \epsilon))$, then $x \mapsto \ln_\epsilon |x|$ is a ϵ^{-1}-Lipschitz function on $\mathbb{R}$. Let $\delta \in (0, 1)$. In this section, we show that a cut-off $\epsilon_n = n^{-\delta}$ on the eigenvalues closest to 0 does not perturb the logarithmic potential at the thermodynamical scale. As we mentioned in Section 7.2 we rely on a technical assumption on $\phi(x)$. Precisely, for any $\delta \in (0, 1)$, we assume that there exists $\eta > 0$ such that for all $t > 0$:

$$\begin{cases} \lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}\left[\left| \frac{1}{n} \sum_{\lambda\in \mathrm{Sp}(\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y}))} \ln|\lambda| \mathbb{1}\{|\lambda| \le n^{-\delta}\} \right| \ge t \right] = -\infty, & \text{(E.16a)} \\ \lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}\left[\int_{|x - t_\phi(\nu_{\mathbf{y}}^m)| \le n^{-\delta}} \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) \ \ln\left|x - t_\phi(\nu_{\mathbf{y}}^m)\right| \le -t \right] = -\infty. & \text{(E.16b)} \end{cases}$$

Physically, this makes explicit that, with large probability, there should not be enough eigenvalues of $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$ around zero so that they contribute macroscopically to the logarithmic potential. This is a consequence the natural fluctuations and repulsion of the eigenvalues of $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$, and we are working to prove it under Definition 7.1 by adapting the arguments of [BABM21a]. Denote $\{\lambda_i\}_{i=1}^{n-1}$ the (sorted) eigenvalues of $\mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})$. We can now state:

Lemma E.7 (*Effect of the cut-off on the expected determinant*)

There exists $\eta > 0$ such that for all $K > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}\left[\left| \frac{1}{n} \ln \mathbb{E} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \frac{1}{n} \ln \mathbb{E}\, e^{\sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} \right| \ge K \right] = -\infty.$$

Proof of Lemma E.7 – We consider η given by eq. (E.16a). Let $t > 0$. We denote $A_t^{(n)}$ the event

$$A_t^{(n)} \equiv \left\{ \left| \frac{1}{n} \sum_{i=1}^{n-1} \ln|\lambda_i| \mathbb{1}\{|\lambda_i| \le n^{-\delta}\} \right| \ge t \right\}.$$

We have for all $\mathbf{y}$ and $t > 0$ ($\bar{A}_t^{(n)}$ being the complementary event to $A_t^{(n)}$):

$$\frac{1}{n} \ln \mathbb{E}_{\mathbf{z}}\, e^{\sum_{i=1}^{n-1} \ln |\lambda_i|} \geq \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}} \Big\{ e^{\sum_{i=1}^{n-1} \ln |\lambda_i|} \mathbb{1}[\bar{A}_t^{(n)}] \Big\} \geq -t + \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}} \Big\{ e^{\sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} \mathbb{1}[\bar{A}_t^{(n)}] \Big\}.$$

So that (using $\ln_{\epsilon_n}(x) \geq \ln(x)$ for all $x > 0$):

$$0 \leq \frac{1}{n} \ln \mathbb{E}\, e^{\sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} - \frac{1}{n} \ln \mathbb{E} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| \leq t - \frac{1}{n} \ln \left[1 - \frac{\mathbb{E}_{\mathbf{z}} \Big\{ e^{\sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} \mathbb{1}[A_t^{(n)}] \Big\}}{\mathbb{E}_{\mathbf{z}} \Big\{ e^{\sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} \Big\}} \right].$$

We know $\ln_{\epsilon_n} |x| \geq -\delta \ln(n)$. By Lemma E.5, for all $\gamma > 0$:

$$\limsup_{n \to \infty} \sup_{\mathbf{y} \in \mathbb{R}^m} \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}} \Big[e^{\gamma \sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} \Big] < +\infty.$$

Fixing $\gamma > 1$ and using Hölder's inequality, there exists therefore $C > 0$ such that for all $K > 0$ and $t \in (0, K)$:

$$\begin{aligned}
&\limsup_{n \to \infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}_{\mathbf{y}} \left[\left| \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \frac{1}{n} \ln \mathbb{E}_{\mathbf{z}}\, e^{\sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|} \right| \geq K \right] \\
&\leq \limsup_{n \to \infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}_{\mathbf{y}} \left[\mathbb{P}_{\mathbf{z}}[A_t^{(n)}]^{1/\gamma} \geq e^{-n(\delta \ln(n) + C)} \Big[1 - e^{n(t-K)} \Big] \right], \\
&\overset{(a)}{\leq} \limsup_{n \to \infty} \frac{1}{n^{1+\eta}} \ln \left\{ \frac{\mathbb{P}[A_t^{(n)}]}{e^{-\gamma n(\delta \ln(n) + C)} [1 - e^{n(t-K)}]^{\gamma}} \right\} \overset{(b)}{=} -\infty,
\end{aligned}$$

in which we used the Markov inequality in (a) and eq. (E.16a) in (b). □ For simplicity we will often abusively denote in the following $\ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| \equiv \sum_{i=1}^{n-1} \ln_{\epsilon_n} |\lambda_i|$ and $\ln_{\epsilon_n} \mathbb{E} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| \equiv \ln \mathbb{E}\, e^{\sum_{i=1}^{n} \ln_{\epsilon_n} |\lambda_i|}$.

Concentration of the logarithmic potential with a cut-off

We show here that discarding the eigenvalues of the Hessian that are close to 0 using a cut-off $\epsilon_n \equiv n^{-\delta}$, we have concentration of the logarithmic potential.

Proposition E.8 (*Concentration of the logarithmic potential*)

Let us fix $\delta < 1/2$ and recall that $\epsilon_n = n^{-\delta}$. Then:

$$\forall t > 0, \ \limsup_{n \to \infty} \sup_{\mathbf{y} \in \mathbb{R}^m} \frac{1}{n^{2(1-\delta)}} \ln \mathbb{P}_{\mathbf{z}} \left[\frac{1}{n} \left| \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \mathbb{E}_{\mathbf{z}} \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| \right| \geq t \right] < 0.$$

Proof of Proposition E.8 – We will try to use traditional Lipschitz concentration bounds (cf e.g. [AGZ10]). We will study under which conditions the function $G(\mathbf{z}) \equiv (1/n) \ln_{\epsilon_n} |\det \tilde{\mathbf{H}}_n^{\mathbf{\Lambda}}(\mathbf{y})|$ is a Lipschitz function of $\mathbf{z} \in \mathbb{R}^{(n-1) \times m}$ (for fixed $\mathbf{y}$). We will do it by bounding $\|\nabla_{\mathbf{z}} G\|_{\infty}$. Let $f_n(x) \equiv \ln_{\epsilon_n} |x|$ for $x \in \mathbb{R}$. We have:

$$\sum_{i=1}^{n-1} \sum_{\mu=1}^{m} \left(\frac{\partial G(\mathbf{z})}{\partial z_{i\mu}} \right)^2 = \frac{1}{n^4} \sum_{i=1}^{n-1} \sum_{\mu=1}^{m} \left[\mathrm{Tr} \Big\{ f_n' \Big(\frac{1}{n} \mathbf{z} \mathbf{\Lambda}(\mathbf{y}) \mathbf{z}^{\intercal} \Big) \Delta_{i\mu} \Big\} \right]^2,$$

in which $\Delta_{i\mu} \in \mathbb{R}^{(n-1)\times(n-1)}$ with $(\Delta_{i\mu})_{jk} \equiv \Lambda_\mu(\mathbf{y})(\delta_{ij} z_{k\mu} + \delta_{ik} z_{j\mu})$. So one shows easily:

$$\sum_{i=1}^{n-1}\sum_{\mu=1}^{m}\left(\frac{\partial G(\mathbf{z})}{\partial z_{i\mu}}\right)^2 = \frac{4}{n^3}\mathrm{Tr}\left[\left(f_n'\Big(\frac{1}{n}\mathbf{z}\boldsymbol{\Lambda}(\mathbf{y})\mathbf{z}^\intercal\Big)\right)^2\left(\frac{1}{n}\mathbf{z}\boldsymbol{\Lambda}(\mathbf{y})^2\mathbf{z}^\intercal\right)\right]. \tag{E.17}$$

Let us recall the Hoffman-Wielandt inequality [HW03]:

Lemma E.9 (*Hoffman-Wielandt inequality for the L_2 norm*)

Let $k \in \mathbb{N}^\star$, and $\mathbf{A}, \mathbf{B} \in \mathcal{S}_k(\mathbb{R})$ be two symmetric matrices with respective eigenvalues $\lambda_1(\mathbf{A}) \leq \cdots \leq \lambda_k(\mathbf{A})$ and $\lambda_1(\mathbf{B}) \leq \cdots \leq \lambda_k(\mathbf{B})$. Then $\sum_{i=1}^k [\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B})]^2 \leq \|\mathbf{A} - \mathbf{B}\|_2^2$.

In particular if $\mathbf{A}$ and $\mathbf{B}$ are positive matrices one has $\mathrm{Tr}[\mathbf{AB}] \leq \sum_i \lambda_i(\mathbf{A})\lambda_i(\mathbf{B})$. We use this in eq. (E.17) along with the n^δ-Lipschizity of f_n:

$$\sum_{i=1}^{n-1}\sum_{\mu=1}^{m}\left(\frac{\partial G(\mathbf{z})}{\partial z_{i\mu}}\right)^2 \leq \frac{4n^{2\delta}}{n^4}\mathrm{Tr}\big[\mathbf{z}\boldsymbol{\Lambda}(\mathbf{y})^2\mathbf{z}^\intercal\big] \leq \frac{4^3 n^{2\delta}\|\mathbf{D}\|_\infty^2}{n^4}\sum_{\mu=1}^{m}\sum_{i=1}^{n-1} z_{i\mu}^2, \tag{E.18}$$

in which we used Lemma E.3. We denote A the event

$$A \equiv \Big\{\frac{1}{n^2}\sum_{\mu=1}^{m}\sum_{i=1}^{n} z_{i\mu}^2 \geq 1 + \alpha\Big\}.$$

It is a classical concentration result (cf e.g. Chapter 3.1 of [Ver18]) that there exists $c > 0$ such that:

$$\mathbb{P}_{\mathbf{z}}\Big[\Big|\sqrt{\frac{1}{n^2}\sum_{\mu,i} z_{\mu i}^2} - \sqrt{\alpha}\Big| \geq t\Big] \leq 2e^{-cn^2t^2}.$$

In particular, this implies

$$\mathbb{P}_{\mathbf{z}}[A] \leq 2e^{-cn^2(\sqrt{1+\alpha}-\sqrt{\alpha})^2}. \tag{E.19}$$

Let us now show that it suffices to prove the bound of Proposition E.8 assuming that A does not occur. Indeed, $n^{-2(1-\delta)}\ln \mathbb{P}_{\mathbf{z}}[A] \leq -cn^{2\delta}$ for a constant $c > 0$, and:

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{\mathbf{y}\in\mathbb{R}^m}\frac{1}{n^{2(1-\delta)}}\ln \mathbb{P}_{\mathbf{z}}\left[\frac{1}{n}\Big|\ln_{\epsilon_n}|\det \mathbf{H}_n^{\boldsymbol{\Lambda}}(\mathbf{y})| - \mathbb{E}_{\mathbf{z}}\ln_{\epsilon_n}|\det \mathbf{H}_n^{\boldsymbol{\Lambda}}(\mathbf{y})|\Big| \geq t\middle| A\right] \\
&\leq \limsup_{n\to\infty}\sup_{\mathbf{y}\in\mathbb{R}^m}\frac{1}{n^{2(1-\delta)}}\ln \mathbb{P}_{\mathbf{z}}\left[\Big|\det \mathbf{H}_n^{\boldsymbol{\Lambda}}(\mathbf{y})|\Big| \geq e^{nt+\mathbb{E}_{\mathbf{z}}\ln_{\epsilon_n}|\det \mathbf{H}_n^{\boldsymbol{\Lambda}}(\mathbf{y})|}\middle| A\right], \\
&\overset{(a)}{\leq} \limsup_{n\to\infty}\sup_{\mathbf{y}\in\mathbb{R}^m}\left\{\frac{1}{n^{2(1-\delta)}}\ln \mathbb{E}_{\mathbf{z}}\Big[|\det \mathbf{H}_n^{\boldsymbol{\Lambda}}(\mathbf{y})|\Big|A\Big] - \frac{1}{n^{2(1-\delta)}}\Big(nt + \mathbb{E}_{\mathbf{z}}\ln_{\epsilon_n}|\det \mathbf{H}_n^{\boldsymbol{\Lambda}}(\mathbf{y})|\Big)\right\}, \\
&\overset{(b)}{<} +\infty.
\end{aligned}$$

We used Markov's inequality in (a). The inequality (b) can be obtained by very similar arguments than the one used to prove Lemma E.5: by rescaling $\mathbf{z}$ by $\|\mathbf{z}\|$, it is easy to see that the event A will not change the scaling of the large deviations of the largest eigenvalue of the Wishart matrix $\mathbf{z}\mathbf{z}^\intercal/n$, so that the bound of Lemma E.5 will also apply when conditioning by the event

A. All in all,

$$\limsup_{n\to\infty} \sup_{\mathbf{y}\in\mathbb{R}^m} \left\{ \frac{1}{n^{2(1-\delta)}} \ln \mathbb{P}_{\mathbf{z}} \left[\frac{1}{n} \left| \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \mathbb{E}_{\mathbf{z}} \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| \right| \geq t \middle| A \right] \mathbb{P}_{\mathbf{z}}[A] \right\} = -\infty.$$

Using the law of total expectation, we can safely ignore the occurrence of A when showing Proposition E.8. Assuming that A is not occurring yields:

$$\sum_{i=1}^{n-1} \sum_{\mu=1}^{m} \left(\frac{\partial G(\mathbf{z})}{\partial z_{i\mu}} \right)^2 \leq \frac{4^3(1+\alpha) n^{2\delta} \|\mathbf{D}\|_\infty^2}{n^2}. \tag{E.20}$$

Recall the Lipschitz concentration of independent variables with laws satisfying the logarithmic Sobolev inequality with a uniform constant c (see for instance [AGZ10] for a proof and an introduction to the logarithmic Sobolev inequalities):

Lemma E.10 (*Herbst*)

Let $n \in \mathbb{N}^\star$ and P be a probability distribution on $\mathbb{R}^n$ satisfying the Logarithmic Sobolev Inequality with constant $c > 0$. Let G be a Lipschitz function on $\mathbb{R}^n$ with Lipschitz constant $\|G\|_{\mathcal{L}}$. Then for all $t > 0$, $\mathbb{P}[|G - \mathbb{E}G| \geq t] \leq 2\exp[-t^2/(2c\|G\|_{\mathcal{L}}^2)]$.

It is easy to check that the Gaussian standard law of $\mathbf{z}$, conditioned by the (extremely probable) event $\overline{A}$, satisfies the Logarithmic Sobolev Inequality with constant $c = 1 + \mathcal{O}_n(1)$. Applying Lemma E.10 alongside eq. (E.20) finishes the proof. □

The logarithmic potential of the asymptotic measure

In this part we relate the expected logarithmic potential to the logarithmic potential of the measure $\mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]$, cf Theorem 7.5.

Proposition E.11 (*Limit of the expected logarithmic potential*)

There exists $\eta > 0$ such that for all $t > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}_{\mathbf{y}} \left[\left| \mathbb{E}_{\mathbf{z}} \frac{1}{n} \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m)) \right| \geq t \right] = -\infty.$$

Proof of Proposition E.11 – The proof goes in two parts. First, we show that there exists $\eta_1 > 0$ such that[1]:

$$\lim_{n\to\infty} \left[n^{\eta_1} \sup_{\mathbf{y}\in\mathbb{R}^m} \left| \mathbb{E}_{\mathbf{z}} \frac{1}{n} \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \int_{\mathbb{R}} \ln_{\epsilon_n} |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) \right| \right] = 0. \tag{E.21}$$

We will then conclude by showing that there exists $\eta_2 > 0$ such that for all $t > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\eta_2}} \mathbb{P}_{\mathbf{y}} \left[\left| \int_{\mathbb{R}} \ln_{\epsilon_n} |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) - \kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m)) \right| \geq t \right] = -\infty. \tag{E.22}$$

[1] Note that this result is uniform over $\mathbf{y}$, and thus stronger than what is needed to show Proposition E.11.

We begin by eq. (E.22). We take η_2 given by eq. (E.16b). We have, since $ln_\epsilon(x) \geq ln(x)$:

$$\begin{aligned}
0 &\geq \int_{\mathbb{R}} \ln_{\epsilon_n} |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) - \kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m)) \\
&= -\delta \ln(n) \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](t_\phi(\nu_{\mathbf{y}}^m) - \epsilon_n, t_\phi(\nu_{\mathbf{y}}^m) + \epsilon_n) - \int_{t_\phi - \epsilon_n}^{t_\phi + \epsilon_n} \ln |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x), \\
&\geq -2 \int_{t_\phi - \epsilon_n}^{t_\phi + \epsilon_n} \ln |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n^{1+\eta_2}} \mathbb{P}\Big[\Big| \int_{\mathbb{R}} \ln_{\epsilon_n} |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) - \kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m)) \Big| \geq t \Big] \\
&\leq \limsup_{n\to\infty} \frac{1}{n^{1+\eta_2}} \mathbb{P}\Big[\int_{t_\phi - \epsilon_n}^{t_\phi + \epsilon_n} \ln |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) \leq -\frac{t}{2} \Big],
\end{aligned}$$

and using eq. (E.16b), we reach eq. (E.22). Let us show eq. (E.21). Its proof is based on the following lemma, a consequence of the analysis of [SB95, BS10]:

Lemma E.12 (*Convergence of the Stieltjes transform*)

Denote $g_n(z)$ the Stieltjes transform of $\mathbf{z}\mathbf{\Lambda}(\mathbf{y})\mathbf{z}^\intercal/n$, and $g_{\alpha,\phi}[\nu_{\mathbf{y}}^m](z)$ the one of $\mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]$, for $z \in \mathbb{C}_+$. Then there exists $\eta \in (0,1)$ such that for all $z \in \mathbb{C}_+$:

$$\lim_{n\to\infty} \{ \sup_{\mathbf{y}\in\mathbb{R}^m} n^\eta |\mathbb{E}_{\mathbf{z}}(g_n(z)) - g_{\alpha,\phi}[\nu_{\mathbf{y}}^m](z)| \} = 0. \tag{E.23}$$

The proof follows quite closely [SB95], using cavity method arguments for random matrices that we used as well in Chapter 5. The interested reader can refer to the Appendix of [MBAB20] for the detailed proof. Let us fix η given by Lemma E.12. As stated for instance in Theorem 2.4.4 of [AGZ10], a consequence of the Stieltjes-Perron inversion (Theorem 1.5) is that for every Borel set $E \subseteq \mathbb{R}$:

$$\lim_{n\to\infty} \sup_{\mathbf{y}\in\mathbb{R}^m} [n^\eta |\mathbb{E}\mu_n(E) - \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](E)|] = 0, \tag{E.24}$$

in which μ_n is the ESD of $\mathbf{z}\mathbf{\Lambda}(\mathbf{y})\mathbf{z}^\intercal/n$. Fix $\eta_1 < \eta$. We have, uniformly over $\mathbf{y}$:

$$\begin{aligned}
&n^{\eta_1} \Big| \mathbb{E}_{\mathbf{z}} \frac{1}{n} \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \int_{\mathbb{R}} \ln_{\epsilon_n} |x - t_\phi(\nu_{\mathbf{y}}^m)| \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x) \Big| \\
&\leq n^{\eta_1} \int_{|x - t_\phi(\nu_{\mathbf{y}}^m)|>1} \ln |x - t_\phi(\nu_{\mathbf{y}}^m)| [\mathbb{E}\mu_n - \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]](\mathrm{d}x) \\
&\quad + \delta \ln(n) n^{\eta_1} \int_{|x - t_\phi(\nu_{\mathbf{y}}^m)|<1} [\mathbb{E}\mu_n - \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]](\mathrm{d}x).
\end{aligned} \tag{E.25}$$

Let us fix $C > 0$ given by item (i) of Lemma E.6. We can bound $t_\phi(\nu_{\mathbf{y}}^m)$ by $||x\phi'||_\infty$. This gives that for n large enough the RHS of eq. (E.25) is bounded (uniformly over $\mathbf{y}$) by:

$$n^{\eta_1} [\ln(C + ||x\phi'||_\infty) + \delta \ln(n)] \{ [\mathbb{E}\mu_n - \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m]](-C, C) \}.$$

Since $\eta_1 < \eta$ we can use eq. (E.24), which shows eq. (E.21). □

Conclusion of the proof

Let us conclude the proof of Lemma 7.10 from all our results of Section E.2.2. We fix $\delta > 0$ such that $\delta < 1/2$. Note that Proposition E.8 is a uniform result on $\mathbf{y}$, much stronger than what we required. Since it shows that the concentration (as a function of $\mathbf{z}$) of the log-determinant is in a scale $n^{1+\epsilon}$ with $\epsilon > 0$, it implies that there exists $\eta > 0$ such that for all $t > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}\Big[\Big|\frac{1}{n}\ln_{\epsilon_n} \mathbb{E}_{\mathbf{z}}|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})| - \frac{1}{n}\mathbb{E}_{\mathbf{z}} \ln_{\epsilon_n} |\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|\Big| \geq t\Big] = -\infty. \tag{E.26}$$

Combining this identity with Lemma E.7 and Proposition E.11, we reach the conclusion of Lemma 7.10.

E.2.3 Proof of Lemma 7.11

Let $\gamma \in (1, \alpha)$. We fix $C > 0$ given by Lemma E.6. Then for all $\mathbf{y} \in \mathbb{R}^m$:

$$\begin{aligned}\kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m)) &\leq \int \mu_{\alpha,\phi}[\nu_{\mathbf{y}}^m](\mathrm{d}x)\ \ln(1+|x|) + \ln(1+|t_\phi(\nu_{\mathbf{y}}^m)|),\\ &\leq \ln(1+\|x\phi'(x)\|_\infty) + \ln(1+C).\end{aligned}$$

By Lemma E.5 we have:

$$\limsup_{n\to\infty} \sup_{\mathbf{y}\in\mathbb{R}^m} \Big[\frac{1}{n}\ln \mathbb{E}|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|\Big] < +\infty.$$

Therefore, in order to prove Lemma 7.11, it only remains to show that

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{E}_{\mathbf{y}}\Big[\exp\Big\{-\frac{\gamma n}{2}\ln\Big(\frac{1}{m}\sum_{\mu=1}^m \phi'(y_\mu)^2\Big)\Big\}\Big] < \infty. \tag{E.27}$$

Let us now prove eq. (E.27). We denote $v \equiv \mathbb{E}_{y\sim\mathcal{N}(0,1)}[\phi'(y)^2]$ and $A \equiv \|\phi'\|_\infty^2$. Since $A < \infty$, we can apply Cramer's theorem to $S \equiv (1/m)\sum_\mu \phi'(y_\mu)^2$, so that we have:

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{E}_{\mathbf{y}}\Big[\exp\Big\{-\frac{\gamma n}{2}\ln\Big(\frac{1}{m}\sum_{\mu=1}^m \phi'(y_\mu)^2\Big)\Big\}\Big] \leq \sup_{S\in(0,A)} [-\frac{\gamma}{2}\ln S - \alpha \mathbf{\Lambda}^\star(S)], \tag{E.28}$$

in which $\mathbf{\Lambda}^\star(S)$ is defined as the Legendre transform of the moment generating function of $\phi'(y)^2$:

$$\mathbf{\Lambda}^\star(S) \equiv \begin{cases} \sup_{\theta\geq 0}\{\theta S - \ln \mathbb{E}_{y\sim\mathcal{N}(0,1)}[e^{\theta\phi'(y)^2}]\} & \text{if } S \geq v,\\ \sup_{\theta\geq 0}\{-\theta S - \ln \mathbb{E}_{y\sim\mathcal{N}(0,1)}[e^{-\theta\phi'(y)^2}]\} & \text{if } S \leq v.\end{cases}$$

By continuity of the involved functions, in order to conclude from eq. (E.28) we just need to be able to show that $(i) : \limsup_{S\uparrow A}(-\mathbf{\Lambda}^\star(S)) < \infty$ and $(ii) : \limsup_{S\downarrow 0}[-\frac{\gamma}{2}\ln S - \alpha\mathbf{\Lambda}^\star(S)] < \infty$. Point (i) is trivial since $\mathbf{\Lambda}^\star(S) \geq 0$ for all $S \in (0, A)$ (it is a rate function). To show (ii), we use the fact that for all $S \in (0, v)$ and $\theta \geq 0$ we have $\mathbf{\Lambda}^\star(S) \geq -\theta S - \ln \mathbb{E}[e^{-\theta\phi'(y)^2}]$. In particular, for $\theta = S^{-1}$ we have $\mathbf{\Lambda}^\star(S) \geq -1 - \ln \mathbb{E}[e^{-S^{-1}\phi'(y)^2}]$. Since $a = \phi'(y)$ has a density φ_a continuous around 0 (Def. 7.1), we fix $a_0 > 0$ such that φ_a is continuous in $[-a_0, a_0]$. For every $\theta > 0$:

$$\ln \mathbb{E}[e^{-\theta\phi'(y)^2}] \leq \ln\Big[\mathbb{E}(e^{-\theta a^2}\mathbb{1}_{|a|\leq a_0}) + e^{-\theta a_0^2}\Big] \leq \ln\Big[(\sup_{|a|\leq a_0}|\varphi_a(a)|)\frac{\sqrt{\pi}}{\sqrt{\theta}} + e^{-\theta a_0^2}\Big],$$

and thus $\ln \mathbb{E}[e^{-\theta\phi'(y)^2}] \leq C-(1/2)\ln\theta$ with a constant $C>0$. Using this bound and the remark before we reach

$$-\frac{\gamma}{2}\ln S - \mathbf{\Lambda}^\star(S) \leq \frac{\alpha-\gamma}{2}\ln S + \alpha(1+C).$$

Since $\alpha-\gamma>0$, we have $\lim_{S\downarrow 0}[-\frac{\gamma}{2}\ln S - \alpha\mathbf{\Lambda}^\star(S)] = -\infty$, which obviously implies point (ii), which in turn shows eq. (E.27).

E.2.4 Proof of eq. (7.20)

Let $t>0$, and fix $\eta>0$ given by Lemma 7.10. We define $E_n^{(t)}$, A_n, B_n:

$$\begin{cases} E_n^{(t)} & \equiv \Big\{\Big|\frac{1}{n}\ln \mathbb{E}_{\mathbf{z}}[|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|] - \kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m))\Big| \geq t\Big\}, \\ A_n & \equiv \frac{1}{n}\ln \mathbb{E}\Big[\mathbb{1}\{L_1(\mathbf{y})\in B\} e^{-\frac{n}{2}\ln(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2)} \mathbb{E}|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|\Big], \\ B_n & \equiv \frac{1}{n}\ln \mathbb{E}\Big[\mathbb{1}\{L_1(\mathbf{y})\in B\} e^{-\frac{n}{2}\ln(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2) + n\kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m))}\Big]. \end{cases} \tag{E.29}$$

A_n is related to the complexity by Lemma 7.8, and by Lemma 7.11 we can apply Varadhan's lemma 1.10 to B_n, which yields

$$\lim_{n\to\infty} B_n = \sup_{\nu\in\mathcal{M}_\phi(B)} [-\frac{1}{2}\mathcal{E}_\phi(\nu) + \kappa_{\alpha,\phi}(\nu, t_\phi(\nu)) - \alpha D_{\mathrm{KL}}(\nu|\mu_G)] \in [-\infty, +\infty). \tag{E.30}$$

The factor α in front of the relative entropy arises as we consider the empirical distribution of m i.i.d. variables. For all $t>0$, we have by definition of A_n, B_n:

$$\begin{cases} A_n - B_n & \geq -t + \frac{1}{n}\ln\Big[1 - \dfrac{\mathbb{E}[\mathbb{1}_{L_1(\mathbf{y})\in B; E_n^{(t)}} e^{-\frac{n}{2}\ln(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2) + n\kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m))}]}{\mathbb{E}[\mathbb{1}_{L_1(\mathbf{y})\in B}\ e^{-\frac{n}{2}\ln(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2) + n\kappa_{\alpha,\phi}(\nu_{\mathbf{y}}^m, t_\phi(\nu_{\mathbf{y}}^m))}]}\Big], \\ A_n - B_n & \leq t + \frac{1}{n}\ln\Big[1 + e^{-nt}\dfrac{\mathbb{E}[\mathbb{1}_{L_1(\mathbf{y})\in B; E_n^{(t)}} e^{-\frac{n}{2}\ln(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2)} \mathbb{E}|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|]}{\mathbb{E}[\mathbb{1}_{L_1(\mathbf{y})\in B} e^{-\frac{n}{2}\ln(\frac{1}{m}\sum_\mu \phi'(y_\mu)^2)} \mathbb{E}|\det \mathbf{H}_n^{\mathbf{\Lambda}}(\mathbf{y})|]}\Big]. \end{cases} \tag{E.31}$$

Using Hölder's inequality and Lemma 7.11, there exists therefore $\gamma>1$ and a constant $C>0$ such that:

$$-t + \frac{1}{n}\ln\Big[1 - \frac{\mathbb{P}[E_n^{(t)}]^{\frac{1}{\gamma}}}{e^{nB_n - nC}}\Big] \leq A_n - B_n \leq t + \frac{1}{n}\ln\Big[1 + \frac{\mathbb{P}[E_n^{(t)}]^{\frac{1}{\gamma}}}{e^{nt + nA_n - nC}}\Big]. \tag{E.32}$$

Assume that $\lim B_n = -\infty$ and $\limsup A_n > -\infty$. Let us fix a lower-bounded sub-sequence $A_{\varphi(n)}$ of A_n, so that $\lim[A_{\varphi(n)} - B_{\varphi(n)}] = +\infty$. However, by eq. (E.32) and Lemma 7.10, we have $\limsup[A_{\varphi(n)} - B_{\varphi(n)}] \leq t$, as $(1/n)\ln\mathbb{P}[E_n^{(t)}] \to -\infty$. So we showed that $\lim B_n = -\infty \Rightarrow \lim A_n = -\infty$, which shows eq. (7.20) in this case.

Let us now assume that $\lim B_n > -\infty$. Using the left inequality of eq. (E.32) and Lemma 7.10, we reach in the same way that $\liminf[A_n - B_n] \geq -t$, which implies that $\liminf A_n > -\infty$. Thus we can use the right inequality of eq. (E.32) to show similarly that $\limsup[A_n - B_n] \leq t$. Taking the $t\to 0$ limit finishes the proof of eq. (7.20).

E.2.5 Annealed and quenched calculations for L_2

We give here a sketch of the generalization of our annealed and quenched calculations to L_2, yielding Theorem 7.6 and Result 7.2. We restrict here to the annealed calculation (the generalization of the quenched calculation is completely similar). The majority of the arguments being identical to the L_1 case, we will only highlight the main differences and give the important intermediary results.

In the Kac-Rice formula, one has to integrate over the overlap $q \equiv \mathbf{x} \cdot \mathbf{x}^\star$ as well. Moreover, we condition over the joint values of $a_\mu \equiv \boldsymbol{\xi}_\mu \cdot \mathbf{x}$ and $b_\mu \equiv (1-q^2)^{-1/2}[(\boldsymbol{\xi}_\mu \cdot \mathbf{x}^\star) - q a_\mu]$, rather than just $\boldsymbol{\xi}_\mu \cdot \mathbf{x}$ (as we did for L_1). Note that (a_μ, b_μ) follows a joint standard Gaussian distribution. Using these definitions we can obtain the counterpart of Lemma 7.8 for L_2:

$$\mathbb{E}\,\mathrm{Crit}_{n,L_2}(B,Q) = \mathcal{C}_n \int_Q \mathrm{d}q\; e^{\frac{n(1+\ln\alpha+\ln(1-q^2))}{2}} \mathbb{E}_{\mathbf{a},\mathbf{b}}[\delta(P_n(\mathbf{a},\mathbf{b}))\mathbb{1}_{L_2(\mathbf{a},\mathbf{b})\in B} e^{-n\mathcal{E}_n(\mathbf{a},\mathbf{b})} \mathbb{E}_{\mathbf{z}}|\det \mathbf{H}_n(\mathbf{a},\mathbf{b})|],$$

in which $\mathcal{C}_n$ is exponentially trivial, and we defined:

$$\begin{cases} P_n(\mathbf{a},\mathbf{b}) & \equiv \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m b_\mu \phi'(a_\mu)[\phi(qa_\mu + \sqrt{1-q^2}b_\mu) - \phi(a_\mu)], \\ \mathcal{E}_n(\mathbf{a},\mathbf{b}) & \equiv \dfrac{1}{2}\ln\Big\{\displaystyle\sum_{\mu=1}^m \phi'(a_\mu)^2[\phi(qa_\mu + \sqrt{1-q^2}b_\mu) - \phi(a_\mu)]^2\Big\}, \\ H_n(\mathbf{a},\mathbf{b}) & \equiv \dfrac{1}{m}\displaystyle\sum_{\mu=1}^m \Big[\phi'(a_\mu)^2 - \theta''(a_\mu)[\phi(\sqrt{1-q^2}b_\mu + qa_\mu) - \phi(a_\mu)]\Big]\mathbf{z}_\mu \mathbf{z}_\mu^\intercal \\ & \quad -\Big(\dfrac{1}{m}\displaystyle\sum_{\mu=1}^m a_\mu \phi'(a_\mu)[\phi(a_\mu) - \phi(qa_\mu + \sqrt{1-q^2}b_\mu)]\Big)\mathrm{I}_{n-2}. \end{cases}$$

Here $\mathbf{z} \in \mathbb{R}^{(n-2)\times m}$ is an i.i.d. standard Gaussian matrix. The condition $P_q(\mathbf{a},\mathbf{b}) = 0$ arises from the conditioning on the nullity of the gradient in the linear subspace of $\{\mathbf{x}\}^\perp$ spanned by $\mathbf{x}^\star$, and $\mathcal{E}_n(\mathbf{a},\mathbf{b})$ from the density of the gradient in the subspace orthogonal to $\{\mathbf{x},\mathbf{x}^\star\}$. A crucial feature of this equation is that the joint distribution of $(L_2(\mathbf{x}), \mathrm{grad}\, L_2(\mathbf{x}), \mathrm{Hess}\, L_2(\mathbf{x}))$ only depends on $\mathbf{x}$ via the overlap $q = \mathbf{x}\cdot\mathbf{x}^\star$ with the "true" solution. Once conditioned over the values of q, it thus becomes clear why the calculations made for L_1 will generalize here.

As in Section 7.3.3, one can then show the concentration of the empirical logarithmic potential on the functional $\kappa_{\alpha,\phi}(q, \nu^m_{\mathbf{a},\mathbf{b}})$, in which $\nu^m_{\mathbf{a},\mathbf{b}} \in \mathcal{M}_1^+(\mathbb{R}^2)$ is now the empirical measure of $\{a_\mu, b_\mu\}_{\mu=1}^m$. We obtain the counterpart of Lemma 7.10: there exists $\eta > 0$ such that for all $t > 0$:

$$\lim_{n\to\infty} \frac{1}{n^{1+\eta}} \ln \mathbb{P}\Big[\Big|\frac{1}{n}\ln \mathbb{E}_{\mathbf{z}}|\det H_n(\mathbf{a},\mathbf{b})| - \kappa_{\alpha,\phi}(q,\nu^m_{\mathbf{a},\mathbf{b}})\Big| \geq t\Big] = -\infty. \tag{E.33}$$

Thanks to this result, we apply then Laplace's method on the overlap q and the empirical measure $\nu \in \mathcal{M}(\mathbb{R}^2)$, using Sanov's theorem 1.9 and Varadhan's lemma 1.10. This yields the result of Theorem 7.6.

As a final note, there exists similar results to the one presented in Section 7.4 that allow to compute the density (and the logarithmic potential) of $\mu_{\alpha,\phi}[q,\nu]$, via the computation of its Stieltjes transform.

E.3 The large deviations in the white Wishart case

In the white Wishart case, we have $\rho(t) = \delta(t-1)$, and the density $\sigma(\lambda)$ is explicitly known, it is the Marchenko-Pastur distribution [MP67]:

$$\sigma(\lambda) = \frac{\alpha}{2\pi}\frac{\sqrt{(\lambda_+(\alpha)-\lambda)(\lambda-\lambda_-(\alpha))}}{\lambda}\mathrm{I}\{\lambda_-(\alpha) < \lambda \leq \lambda_+(\alpha)\}, \tag{E.34}$$

with $\lambda_\pm(\alpha) \equiv (1 \pm \alpha^{-1/2})^2$. One can also explicitly solve the Marchenko-Pastur equation (8.1) (it is just a quadratic equation in this case) and obtains for $x \geq \lambda_+(\alpha)$:

$$\begin{cases} G_\sigma(x) &= \dfrac{1-\alpha+\alpha x - \alpha\sqrt{(x-\lambda_+(\alpha))(x-\lambda_-(\alpha))}}{2x}, \\ \overline{G}_\sigma(x) &= \dfrac{1-\alpha+\alpha x + \alpha\sqrt{(x-\lambda_+(\alpha))(x-\lambda_-(\alpha))}}{2x}. \end{cases}$$

By Result 8.1, this implies that the rate function $I(x)$ satisfies for every $x \geq \lambda_+(\alpha)$:

$$I(x) = \frac{\alpha\beta}{2}\int_{\lambda_+(\alpha)}^{x} \frac{\sqrt{(u-\lambda_+(\alpha))(u-\lambda_-(\alpha))}}{u}\,\mathrm{d}u. \tag{E.35}$$

On the other hand, direct calculations using the joint law of eigenvalues of a Wishart matrix give the following expression of the rate function (see e.g. Theorem 2.4 of [BG20]) for $x \geq \lambda_+(\alpha)$:

$$I(x) = \beta\Big\{\frac{\alpha x}{2} - \frac{\alpha-1}{2}\ln x - \int \mathrm{d}\lambda\,\sigma(\lambda)\ln(x-\lambda) - \frac{1}{2}\Big[1 + \frac{1}{\alpha} + \ln\alpha\Big]\Big\}. \tag{E.36}$$

The logarithmic potential of the Marchenko-Pastur law is known analytically, as stated in Proposition II.1.5 of [Far14]. More precisely, we have for all $x \geq \lambda_+(\alpha)$:

$$\begin{aligned}&\int \mathrm{d}\lambda\,\sigma(\lambda)\ln(x-\lambda) \\ &\quad= \frac{\alpha x}{2} - \frac{\alpha-1}{2}\ln x - \frac{1}{2}\Big[1+\frac{1}{\alpha}+\ln\alpha\Big] - \frac{\alpha}{2}\int_{\lambda_+(\alpha)}^{x}\frac{\sqrt{(u-\lambda_+(\alpha))(u-\lambda_-(\alpha))}}{u}\,\mathrm{d}u.\end{aligned}$$

It is then immediate to see that eq. (E.35) and eq. (E.36) are equivalent, validating Result 8.1 in this simple (yet important) case.

E.4 The phase transition in the rate function

In this section, we investigate possible discontinuities in the derivatives of the rate function $I(x)$, when $d_{\max} > 0$ and $x_c(\rho)$ is finite. In this case, the function $\overline{G}_\sigma(x)$ is constant and equal to $\alpha/d_{\max}$ for $x \geq x_c(\rho)$. Recall that if $s_{\max} \leq x \leq x_c(\rho)$, $\overline{G}_\sigma(x)$ is the second branch to the Marchenko-Pastur equation (8.1). This equation can be written as $F_\sigma(G) = x$, with

$$F_\sigma(G) = \frac{1}{G} + \alpha\int \mathrm{d}t\rho(t)\frac{t}{\alpha - tG}. \tag{E.37}$$

By differentiating the relation $F_\sigma(\overline{G}_\sigma(x)) = x$, we find

$$\overline{G}'_\sigma(x) = 1/F'_\sigma(\overline{G}_\sigma(x)).$$

Let us assume that $\rho(t) \sim (d_{\max} - t)^\eta$ with $\eta > 0$ and t close to $d_{\max}$.

If $\eta \geq 1$, we have $G'_\rho(d_{\max}) < \infty$, so that $F'_\sigma(\alpha/d_{\max}) < \infty$, and thus $\overline{G}'_\sigma(x) \to 1/F'_\sigma(\alpha/d_{\max}) > 0$ as $x \uparrow x_c(\rho)$. The transition in $I(x)$ is thus of second order in this case, as $\overline{G}'_\sigma(x)$ is discontinuous.

If we now assume that $\eta < 1$, we have $G'_\rho(d_{\max}) = +\infty$. By eq. (E.37), this implies that $\overline{G}'_\sigma(x) \to 0$ as $x \uparrow x_c(\rho)$. Thus in this case both $\overline{G}_\sigma$ and $\overline{G}'_\sigma$ are continuous in $x = x_c(\rho)$. We can differentiate the relation $F_\sigma(\overline{G}_\sigma(x)) = x$ once more, and we find easily:

$$\overline{G}''_\sigma(x) = -\frac{F''_\sigma(\overline{G}_\sigma(x))}{F'_\sigma(\overline{G}_\sigma(x))^3}. \tag{E.38}$$

From eqs. (E.37), (E.38), one can show that $\overline{G}''_\sigma(x) \to 0$ as $x \uparrow x_c(\rho)$ if and only if $\eta < 1/2$. In particular, for any $1/2 \leq \eta < 1$, the transition in $I(x)$ is of third order.

Differentiating three times, one can show in a similar way that the transition is of fourth order if and only if $\eta \in [1/3, 1/2)$. Generalizing this to any order, we conjecture that $I(x)$ is smooth at any point $x \neq x_c(\rho)$, and that the first discontinuous derivative of the rate function at $x = x_c(\rho)$ is $I^{(k+1)}(x)$, with $\eta \in [1/k, 1/(k-1))$ (with the convention $1/0 = +\infty$).

E.5 Technicalities on spherical integrals

E.5.1 Simplifying $J_1(\theta, x)$

In this section, we simplify the expression of $J_1(\theta, x)$ when $\theta \leq \theta_c(x) \equiv G_\sigma(x)$. We start from eq. (8.12):

$$J_1(\theta, x) = \inf_{\gamma > \theta x} \left[\frac{\gamma}{2} - \frac{1}{2}\int \mathrm{d}u\, \sigma(u) \ln(\gamma - \theta u)\right] - \frac{1}{2}. \tag{E.39}$$

When $\theta \leq G_\sigma(x)$ the infimum is reached in $\gamma^\star = \theta G_\sigma^{-1}(\theta)$. This implies

$$J_1(\theta, x) = \frac{\theta G_\sigma^{-1}(\theta)}{2} - \frac{1}{2}\ln\theta - \frac{1}{2}\int \mathrm{d}u\, \sigma(u) \ln(G_\sigma^{-1}(\theta) - u)\big] - \frac{1}{2}.$$

Let us differentiate this expression with respect to θ:

$$\partial_\theta J_1(\theta, x) = \frac{G_\sigma^{-1}(\theta)}{2} + \frac{\theta}{2G'_\sigma(G_\sigma^{-1}(\theta))} - \frac{1}{2\theta} - \frac{G_\sigma(G_\sigma^{-1}(\theta))}{2G'_\sigma(G_\sigma^{-1}(\theta))} = \frac{G_\sigma^{-1}(\theta)}{2} - \frac{1}{2\theta}.$$

Using now the Marchenko-Pastur equation (8.1) we can simplify this into:

$$\partial_\theta J_1(\theta, x) = \alpha \int \mathrm{d}t\, \rho(t)\, \frac{t}{\alpha - t\theta} = F'_1(\theta).$$

Since $J_1(0, x) = F_1(0) = 0$, this implies that for every $\theta \leq \theta_c(x)$ we have $J_1(\theta, x) = F_1(\theta)$, which justifies the claim made in the main text.

E.5.2 Derivations of $F_2(\theta)$ and $J_2(\theta, x)$

The derivation of $F_2(\theta)$

We start from the definition of $F_2(\theta)$:

$$F_2(\theta) = \lim_{n\to\infty} \left\{ \frac{1}{n} \ln \int \mathcal{D}\mathbf{z} \int_{\|\mathbf{e}\|^2=1} \mathrm{d}\mathbf{e} \int_{\|\mathbf{f}\|^2=1} \mathrm{d}\mathbf{f}\; e^{\frac{\theta n}{\sqrt{m}} \sum_{i,\mu} \sqrt{d_\mu} e_i z_{\mu i} f_\mu} \right\}.$$

Integrating over $\mathbf{z}$ yields:

$$F_2(\theta) = \lim_{n\to\infty}\left\{\frac{1}{n}\ln\int_{\|\mathbf{f}\|^2=1} \mathrm{d}\mathbf{f}\, e^{\frac{\theta^2 n^2}{2m}\sum_\mu d_\mu f_\mu^2}\right\} = \lim_{n\to\infty}\left\{\frac{1}{n}\ln\frac{\int \mathrm{d}\mathbf{f}\,\delta(||\mathbf{f}||^2-m)\; e^{\frac{\theta^2 n^2}{2m^2}\sum_\mu d_\mu f_\mu^2}}{\int \mathrm{d}\mathbf{f}\,\delta(||\mathbf{f}||^2-m)}\right\}.$$

We introduce a Lagrange multiplier γ to fix the norm of $\mathbf{f}$. This yields (recall $\alpha = m/n$):

$$F_2(\theta) = \inf_{\gamma\geq\theta^2 d_{\max}/\alpha^2}\left\{\frac{\alpha\gamma}{2} - \frac{1}{2n}\ln\det\left(\gamma \mathrm{I}_m - \frac{\theta^2}{\alpha^2}\mathbf{D}_m\right)\right\} - \frac{\alpha}{2} + \mathcal{O}_n(1).$$

The condition $\gamma \geq \theta^2 d_{\max}/\alpha^2$ arises as the matrix inside the determinant must be positive. Changing variables by letting $\gamma = \theta^2\gamma'/\alpha^2$ we arrive at:

$$F_2(\theta) = \frac{\alpha}{2}\inf_{\gamma\geq d_{\max}}\left[\frac{\theta^2\gamma}{\alpha^2} - \int \mathrm{d}t\,\rho(t)\,\ln(\gamma - t)\right] - \frac{\alpha}{2}\ln\frac{\theta^2}{\alpha^2} - \frac{\alpha}{2}.$$

This ends the derivation of the expression of $F_2(\theta)$ given in the main text.

Computing $J_2(\theta, x)$

The goal of this section is to compute $J_2(\theta, x)$. More precisely, we will first show eq. (E.42), which will then be simplified, precisely showing the transition phenomenon described in the main text.

We start from the definition of $J_2(\theta, x)$ (we omit the writing of the $n \to \infty$ limit):

$$\begin{aligned} J_2(\theta, x) &= \frac{1}{n}\ln\int_{\|\mathbf{e}\|^2=1}\mathrm{d}\mathbf{e}\int_{\|\mathbf{f}\|^2=1}\mathrm{d}\mathbf{f}\;\exp\left\{\frac{\theta n}{\sqrt{m}}\sum_{i,\mu}\sqrt{d_\mu}e_i z_{\mu i} f_\mu\right\}, \\ &= \frac{1}{n}\ln\frac{\int \mathrm{d}\mathbf{e}\int\mathrm{d}\mathbf{f}\;\delta(||\mathbf{e}||^2-n)\;\delta(||\mathbf{f}^2||-m)\;\exp\left\{\theta\frac{\sqrt{n}}{m}\sum_{i,\mu}\sqrt{d_\mu}e_i z_{\mu i} f_\mu\right\}}{\int \mathrm{d}\mathbf{e}\int\mathrm{d}\mathbf{f}\;\delta(||\mathbf{e}||^2-n)\;\delta(||\mathbf{f}^2||-m)}. \end{aligned}$$

We introduce two Lagrange multipliers to fix the norms of $\mathbf{e}$ and $\mathbf{f}$. Let us start with the computation of the denominator:

$$\begin{aligned} &\frac{1}{n}\ln\int\mathrm{d}\mathbf{e}\int\mathrm{d}\mathbf{f}\,\delta(||\mathbf{e}||^2-n)\delta(||\mathbf{f}^2||-m) \\ &\qquad\simeq \inf_{\Lambda_e,\Lambda_f\geq 0}\left[\frac{\Lambda_e}{2} + \frac{\alpha\Lambda_f}{2} - \frac{1}{2}\ln\Lambda_e - \frac{\alpha}{2}\ln\Lambda_f + \frac{(1+\alpha)}{2}\ln 2\pi\right]. \end{aligned}$$

The positivity constraint on Λ_e, Λ_f arises naturally for the Gaussian integral to be well-defined. This is easily solved by $\Lambda_e = \Lambda_f = 1$, and we arrive at:

$$\frac{1}{n}\ln\int\mathrm{d}\mathbf{e}\int\mathrm{d}\mathbf{f}\;\delta(||\mathbf{e}||^2-n)\;\delta(||\mathbf{f}^2||-m) \simeq \frac{(1+\alpha)}{2}(1+\ln 2\pi). \tag{E.40}$$

We use the same method to compute the numerator:

$$\begin{aligned} &\frac{1}{n}\ln\int\mathrm{d}\mathbf{e}\int\mathrm{d}\mathbf{f}\;\delta(||\mathbf{e}||^2-n)\;\delta(||\mathbf{f}^2||-m)\;\exp\left\{\theta\frac{\sqrt{n}}{m}\sum_{i,\mu}\sqrt{d_\mu}e_i z_{\mu i} f_\mu\right\} \\ &\simeq \inf_{\Lambda_e,\Lambda_f\geq 0}\left[\frac{\Lambda_e}{2} + \frac{\alpha\Lambda_f}{2} - \frac{1}{2n}\ln\det\begin{pmatrix}\Lambda_e\mathrm{I}_n & \frac{\theta}{\sqrt{\alpha}}\frac{\mathbf{z}^\intercal}{\sqrt{m}}\sqrt{\mathbf{D}_m} \\ \frac{\theta}{\sqrt{\alpha}}\sqrt{\mathbf{D}_m}\frac{\mathbf{z}}{\sqrt{m}} & \Lambda_f\mathrm{I}_m\end{pmatrix} + \frac{(1+\alpha)}{2}\ln 2\pi\right]. \end{aligned}$$

We can compute the determinant of the block matrix easily, and we arrive at:

$$\frac{1}{n}\ln\int \mathrm{d}\mathbf{e}\int \mathrm{d}\mathbf{f}\;\delta(||\mathbf{e}||^2-n)\;\delta(||\mathbf{f}^2||-m)\;\exp\Big\{\theta\frac{\sqrt{n}}{m}\sum_{i,\mu}\sqrt{d_\mu}e_i z_{\mu i} f_\mu\Big\}$$
$$\simeq \inf_{\Lambda_e,\Lambda_f\geq 0}\Big[\frac{\Lambda_e}{2}+\frac{\alpha\Lambda_f}{2}-\frac{\alpha-1}{2}\ln\Lambda_f-\frac{1}{2n}\ln\det\left(\Lambda_e\Lambda_f \mathrm{I}_n-\frac{\theta^2}{\alpha}\mathbf{H}_n\right)+\frac{(1+\alpha)}{2}\ln 2\pi\Big].$$

Note that the matrix inside the log-det must be positive, which constrains $\Lambda_e\Lambda_f \geq \theta^2 x/\alpha$, as we assumed $\lambda_{\max}(\mathbf{H}_n)\simeq x$. All in all, we have, taking $n\to\infty$:

$$\frac{1}{n}\ln\int \mathrm{d}\mathbf{e}\int \mathrm{d}\mathbf{f}\;\delta(||\mathbf{e}||^2-n)\;\delta(||\mathbf{f}^2||-m)\;\exp\Big\{\theta\frac{\sqrt{n}}{m}\sum_{i,\mu}\sqrt{d_\mu}e_i z_{\mu i} f_\mu\Big\} \tag{E.41}$$
$$\simeq \inf_{\substack{\Lambda_e,\Lambda_f\geq 0\\ \text{s.t. } \alpha\Lambda_e\Lambda_f\geq\theta^2 x}}\Big[\frac{\Lambda_e}{2}+\frac{\alpha\Lambda_f}{2}-\frac{\alpha-1}{2}\ln\Lambda_f-\frac{1}{2}\int \mathrm{d}\lambda\,\sigma(\lambda)\,\ln\left(\Lambda_e\Lambda_f-\frac{\theta^2}{\alpha}\lambda\right)+\frac{(1+\alpha)}{2}\ln 2\pi\Big].$$

Combining eqs. (E.40) and (E.41) yields the general formula for J_2:

$$J_2(\theta,x) \tag{E.42}$$
$$=\frac{1}{2}\inf_{\substack{\Lambda_e,\Lambda_f\geq 0\\ \text{s.t. } \alpha\Lambda_e\Lambda_f\geq\theta^2 x}}\Big[\Lambda_e+\alpha\Lambda_f-(\alpha-1)\ln\Lambda_f-\int \mathrm{d}\lambda\,\sigma(\lambda)\ln\left(\Lambda_e\Lambda_f-\frac{\theta^2}{\alpha}\lambda\right)\Big]-\frac{1+\alpha}{2}.$$

The transition in $J_2(\theta,x)$

We start from the expression of $J_2(\theta,x)$ of eq. (E.42). The variational parameters Λ_e,Λ_f can saturate, which is associated to a phase transition. At this point $J_2(\theta,x)$ will become sensitive to the largest eigenvalue of $\mathbf{H}_n$ (assumed to be equal to x).

Given the infimum in eq. (E.42), this phase transition occurs for $\theta=\theta_c(x)$ such that the corresponding values of Λ_e,Λ_f satisfy $\alpha\Lambda_e\Lambda_f=\theta_c(x)^2 x$. From this equation and the equations on Λ_e,Λ_f obtained by making the derivative inside the infimum equal to 0 (which is valid for $\theta\leq\theta_c(x)$), it is easy to obtain

$$\theta_c(x)=\sqrt{xG_\sigma(x)^2+(\alpha-1)G_\sigma(x)}.$$

The case $\theta\leq\theta_c(x)$ – In this case $J_2(\theta,x)$ is not sensitive to the value of x, and we can use a very useful expression derived in Chapter 7 (more precisely eq. (7.23)) for the log-potential of $\sigma(\lambda)$. For any $x\geq s_{\max}$:

$$\int \mathrm{d}\lambda\,\sigma(\lambda)\ln(x-\lambda)=\inf_{0<g<G_\sigma(s_{\max})}\Big[-\ln g+zg+\alpha\int \mathrm{d}t\,\rho(t)\ln(\alpha-tg)\Big]-1-\alpha\ln\alpha.$$

This infimum is attained at $g=G_\sigma(x)$, as it is the unique zero of the derivative of the expression above in the interval $(0,G_\sigma(s_{\max}))$, by eq. (8.1). We can then write eq. (E.42) as:

$$J_2(\theta,x)=-\frac{\alpha(1-\ln\alpha)}{2}-\frac{1}{2}\ln\frac{\theta^2}{\alpha}+\frac{1}{2}\inf_{0<g<G_\sigma(s_{\max})}\inf_{\substack{\Lambda_e,\Lambda_f\geq 0\\ (\alpha\Lambda_e\Lambda_f\geq\theta^2 x)}}\Big[\Lambda_e+\alpha\Lambda_f \tag{E.43}$$
$$-(\alpha-1)\ln\Lambda_f+\ln g-\frac{\alpha\Lambda_e\Lambda_f}{\theta^2}g-\alpha\int \mathrm{d}t\,\rho(t)\ln(\alpha-tg)\Big].$$

Since we are in the "no-saturation" regime, we can use the zero-gradient equations on Λ_e, Λ_f:

$$\begin{cases} \Lambda_f &= \theta^2/(\alpha g), \\ \Lambda_e &= \theta^2/g - (\alpha - 1). \end{cases}$$

Plugging this back into eq. (E.43) we obtain:

$$J_2(\theta, x) = \frac{1}{2} \inf_{0<g<G_\sigma(s_{\max})} \Big[\frac{\theta^2}{g} + \alpha \ln g - \alpha \ln \frac{\theta^2}{\alpha} - \alpha \int \mathrm{d}t\, \rho(t) \ln(\alpha - tg) + \alpha(\ln \alpha - 1) \Big].$$

Changing variables $\gamma = \alpha/g$, we reach:

$$J_2(\theta, x) = \frac{\alpha}{2} \inf_{\gamma \geq \alpha/G_\sigma(s_{\max})} \Big[\frac{\gamma \theta^2}{\alpha^2} - \int \mathrm{d}t \rho(t) \ln(\gamma - t) \Big] - \frac{\alpha}{2} \Big(1 + \ln \frac{\theta^2}{\alpha^2} \Big). \tag{E.44}$$

In order to map $J_2(\theta, x)$ to $F_2(\theta)$, we must only show that the Lagrange multiplier γ in eq. (E.44) does not "saturate" for $\theta \leq \theta_c(x)$. This is easily shown using the Marchenko-Pastur equation (8.1). Indeed since $\theta \leq \theta_c(x)$ we have $\theta \leq \theta_c(s_{\max})$, and thus:

$$\begin{aligned} \theta^2 &\leq s_{\max} G_\sigma(s_{\max})^2 + (\alpha - 1) G_\sigma(s_{\max}) \leq \alpha G_\sigma(s_{\max}) \Big[1 + G_\sigma(s_{\max}) \Big(\frac{s_{\max}}{\alpha} - \frac{1}{\alpha G_\sigma(s_{\max})} \Big) \Big], \\ &\leq \alpha G_\sigma(s_{\max}) \Big[1 + G_\sigma(s_{\max}) \int \frac{\mathrm{d}t\, \rho(t)\, t}{\alpha - t G_\sigma(s_{\max})} \Big] \leq \alpha^2 \int \frac{\mathrm{d}t\, \rho(t)}{\alpha/G_\sigma(s_{\max}) - t}. \end{aligned}$$

This precisely means that the infimum in eq. (E.44) will be attained for a point γ which is a critical point of the functional inside the infimum:

$$\frac{\theta^2}{\alpha^2} = \int \mathrm{d}t\, \rho(t) \frac{1}{\gamma - t},$$

i.e. there is no saturation in both J_2 and F_2, and therefore we have $J_2(\theta, x) = F_2(\theta)$ in this case.

The case $\theta \geq \theta_c(x)$ – In this case, we have a "saturation" in the infimum of eq. (E.42). More precisely, the Λ_e, Λ_f attaining the infimum satisfy $\alpha \Lambda_e \Lambda_f = \theta^2 x$. One can solve the infimum over Λ_e, Λ_f constrained by this equality. Introducing a Lagrange parameter ρ, we reach:

$$\begin{aligned} J_2(\theta, x) = &- \frac{1 + \alpha}{2} - \frac{1}{2} \ln \frac{\theta^2}{\alpha} - \frac{1}{2} \int \mathrm{d}\lambda\, \sigma(\lambda) \ln(x - \lambda) \\ &+ \frac{1}{2} \inf_{\Lambda_e, \Lambda_f \geq 0} \operatorname*{extr}_{\rho} \Big[\Lambda_e + \alpha \Lambda_f - (\alpha - 1) \ln \Lambda_f - \rho \Big(\Lambda_e \Lambda_f - \frac{\theta^2 x}{\alpha} \Big) \Big]. \end{aligned}$$

The extr notation denotes solving the associated zero-gradient equation, as is standard with Lagrange multipliers. One can now solve the infimum over Λ_e, Λ_f easily, and we reach:

$$J_2(\theta, x) = \frac{1}{2} \operatorname*{extr}_{\rho} \Big[\frac{\alpha}{\rho} + (\alpha - 1) \ln \rho + \frac{\rho \theta^2 x}{\alpha} - \ln \frac{\theta^2}{\alpha} - \int \mathrm{d}\lambda\, \sigma(\lambda) \ln(x - \lambda) \Big] - \frac{1 + \alpha}{2}.$$

This can also be solved easily, and finally we have, for $\theta \geq \theta_c(x)$:

$$\begin{aligned} J_2(\theta, x) = \frac{1}{2} \Big[&- (1 + \alpha) - \alpha \ln \frac{\theta^2}{\alpha} - (\alpha - 1) \ln(2x) + \sqrt{(\alpha - 1)^2 + 4x\theta^2} \\ &+ (\alpha - 1) \ln \left[1 - \alpha + \sqrt{(\alpha - 1)^2 + 4x\theta^2} \right] - \int \mathrm{d}\lambda\, \sigma(\lambda) \ln(x - \lambda) \Big]. \end{aligned}$$

This ends the argument by justifying all the expressions given for $J_2(\theta, x)$ in the main text.

A remark on the case $d_{\max} \leq 0$ – In this case $s_{\max} \leq 0$, and the transition we described does not take place, as $\Lambda_e, \Lambda_f \geq 0$ can not satisfy $\alpha\Lambda_e\Lambda_f = \theta^2 x < 0$. The difference between the quenched and annealed integrals in this case has, as far as we know, not been investigated before, and it remains an open question. Importantly, in this setting the first tilting allowed to derive the large deviations, as emphasized in the main text, so that solving this question is not crucial for our purpose.

E.6 Simplifying the rate function

E.6.1 When $x < x_{\max}$, in the first tilting

The goal of this section is to show, for all $x < x_{\max}$:

$$I(x) = \sup_{\theta \in (0,\theta_{\max})} [J_1(\theta, x) - F_1(\theta)] = \frac{1}{2}\int_{s_{\max}}^{x} [\overline{G}_\sigma(u) - G_\sigma(u)]\mathrm{d}u,$$

and that the maximum in θ is reached in $\theta_x = \overline{G}_\sigma(x)$. Recall eq. (8.18):

$$J_1(\theta, x) = \begin{cases} F_1(\theta) = -\dfrac{\alpha}{2}\displaystyle\int \mathrm{d}t\rho(t)\ln(1-\alpha^{-1}\theta t) & \text{if } \theta \leq G_\sigma(x), \\ \dfrac{\theta x - 1 - \ln\theta}{2} - \dfrac{1}{2}\displaystyle\int \mathrm{d}\lambda\ \sigma(\lambda)\ln(x-\lambda) & \text{if } \theta > G_\sigma(x). \end{cases}$$

Differentiating with respect to θ, we reach:

$$\partial_\theta[J_1(\theta, x) - F_1(\theta)] = \begin{cases} 0 & \text{if } \theta \leq G_\sigma(x), \\ \dfrac{\theta x - 1}{2\theta} - \dfrac{\alpha}{2}\displaystyle\int \mathrm{d}t\ \rho(t)\frac{t}{\alpha - \theta t} & \text{if } \theta > G_\sigma(x). \end{cases}$$

So the supremum $\sup_{\theta\in(0,\theta_{\max})}[J_1(\theta, x) - F_1(\theta)]$ is attained for $\theta = \theta_x > G_\sigma(x)$ that satisfies:

$$x = \frac{1}{\theta} + \alpha\int \mathrm{d}t\rho(t)\frac{t}{\alpha - \theta t}.$$

Note that this is exactly the Marchenko-Pastur equation (8.1), so that θ_x is precisely the second "branch" $\theta_x = \overline{G}_\sigma(x)$. Moreover, we know that $I(s_{\max}) = 0$, and we conclude by noticing that:

$$I'(x) = \partial_x[J_1(\theta_x, x) - F_1(\theta_x)] = \frac{\theta_x}{2} - \frac{1}{2}\int \frac{\mathrm{d}\lambda\,\sigma(\lambda)}{x-\lambda} = \frac{1}{2}[\overline{G}_\sigma(x) - G_\sigma(x)].$$

E.6.2 The second tilting

Our objective is to show, for all $x \geq s_{\max}$:

$$I(x) = \sup_{\theta\geq 0}[J_2(\theta, x) - F_2(\theta)] \stackrel{?}{=} \frac{1}{2}\int_{s_{\max}}^{x} [\overline{G}_\sigma(u) - G_\sigma(u)]\mathrm{d}u. \tag{E.45}$$

Recall the functions J_2 and F_2 (with $\theta_c(x) = \sqrt{xG_\sigma(x)^2 + (\alpha-1)G_\sigma(x)}$):

$$J_2(\theta,x) = \begin{cases} F_2(\theta) = \dfrac{\alpha}{2} \inf\limits_{\gamma \geq d_{\max}} \Big[\dfrac{\gamma\theta^2}{\alpha^2} - \displaystyle\int \mathrm{d}t\rho(t) \ln(\gamma - t) - 1 - \ln \dfrac{\theta^2}{\alpha^2} \Big] & \text{if } \theta \leq \theta_c(x), \\ \dfrac{\alpha-1}{2} \ln \Big[\dfrac{1-\alpha+\sqrt{(\alpha-1)^2+4x\theta^2}}{2x} \Big] - \dfrac{1+\alpha}{2} - \dfrac{\alpha}{2} \ln \dfrac{\theta^2}{\alpha} & \\ \qquad + \dfrac{1}{2}\sqrt{(\alpha-1)^2+4x\theta^2} - \dfrac{1}{2} \displaystyle\int \mathrm{d}\lambda\, \sigma(\lambda) \ln(x-\lambda) & \text{if } \theta \geq \theta_c(x). \end{cases}$$

We perform the change of variable $\theta(\tau,x)^2 \equiv x\tau^2 + (\alpha-1)\tau$. At the critical value $\theta_c(x)$, we have $\tau_c(x) = G_\sigma(x)$. We obtain the expression of the rate function as $I(x) = \sup_{\tau \geq G_\sigma(x)} I(x,\tau)$, with $I(x,\tau) = J(\tau,x) - F(\tau,x)$, in which we naturally defined:

$$J(\tau,x) \equiv \frac{1}{2}\Big\{ -2 - \alpha \ln \Big[\frac{\tau x}{\alpha} + 1 - \frac{1}{\alpha} \Big] - \ln(\tau) + 2x\tau - \int \mathrm{d}\lambda\, \sigma(\lambda) \ln(x-\lambda) \Big\}.$$

Similarly, we have the following expression for $F(\tau,x)$:

$$F(\tau,x) \equiv \begin{cases} \dfrac{\alpha}{2} \displaystyle\int_0^{\theta(\tau,x)^2/\alpha^2} [G_\rho^{-1}(u) - \frac{1}{u}] \mathrm{d}u & \text{if } \theta(\tau,x)^2 \leq \alpha^2 G_\rho(d_{\max}), \\ \dfrac{\alpha}{2} \Big[\dfrac{d_{\max}\theta(\tau,x)^2}{\alpha^2} - \displaystyle\int \mathrm{d}t\rho(t) \ln(d_{\max} - t) - 1 - \ln \dfrac{\theta(\tau,x)^2}{\alpha^2} \Big] & \text{if } \theta(\tau,x)^2 \geq \alpha^2 G_\rho(d_{\max}). \end{cases}$$

Using these expressions for J and F, we then compute $\tau_x \equiv \arg\max_{\tau \geq G_\sigma(x)} [J(\tau,x) - F(\tau,x)]$:

$$\partial_\tau [J(\tau,x) - F(\tau,x)] = \tag{E.46}$$
$$\begin{cases} \dfrac{(2\tau x + \alpha - 1)}{2\alpha\tau} (\alpha - \tau G_\rho^{-1}[\theta(\tau,x)^2/\alpha^2]) & \text{if } \theta_c(x)^2 \leq \theta(\tau,x)^2 \leq \alpha^2 G_\rho(d_{\max}), \\ \dfrac{(2\tau x + \alpha - 1)}{2\alpha\tau} (\alpha - \tau d_{\max}) & \text{if } \theta(\tau,x)^2 \geq \alpha^2 G_\rho(d_{\max}). \end{cases}$$

For all $s_{\max} \leq x \leq x_c(\rho) \equiv d_{\max} G_\rho(d_{\max})^2 + (\alpha^{-1} - 1) d_{\max}$, the equation $\alpha = \tau G_\rho^{-1}[\theta(\tau,x)^2/\alpha^2]$ is again the Marchenko-Pastur equation (8.1), with $\omega = \tau$. Since $\tau_x > G_\sigma(x)$, it is easy to check from eq. (E.46) that the supremum must be attained in $\tau_x = \overline{G}_\sigma(x)$. This is true even if $x > x_c(\rho)$, as then the maximum is attained in $\tau = \alpha/d_{\max} = \overline{G}_\sigma(x)$, again from eq. (E.46). Moreover we can compute (from the first equality in eq. (E.45)):

$$I'(x) = \partial_x (J-F)(\tau_x, x) + (\partial_x \tau_x) \underbrace{\partial_\tau (J-F)(\tau_x,x)}_{=0} = \partial_x [J-F](\tau_x,x) = \frac{1}{2}[\tau_x - G_\sigma(x)].$$

Since $I(s_{\max}) = 0$ and $\tau_x = \overline{G}_\sigma(x)$, this implies eq. (E.45).

RÉSUMÉ

Le déluge croissant de données qui a rythmé la dernière décennie a donné naissance à des techniques modernes dans le domaine de l'intelligence artificielle. Ces méthodes sont basées sur l'optimisation d'un très grand nombre de paramètres par l'exploitation d'une quantité gargantuesque de données, et ces algorithmes sont désormais l'état de l'art pour des tâches aussi diverses que la classification d'images, le traitement automatique des langues, ou la reconnaissance vocale, et leurs performances excèdent régulièrement les capacités humaines. En conséquence, de nombreuses recherches se sont concentrées sur la construction d'une théorie mathématique qui pourrait expliquer l'efficacité de ces algorithmes, créant un fort gain d'intérêt pour les statistiques en haute dimension, où la quantité de données et le nombre de paramètres sont tous deux très grands. Nous analysons ici quelques pièces de cet immense puzzle à travers le prisme de la physique statistique, en empruntant également aux probabilités et à la théorie des matrices aléatoires. Ces outils nous permettent de proposer trois approches au problème de l'apprentissage statistique en haute dimension. Dans la première nous revisitons un classique de la physique statistique, les expansions de haute température. Nous expliquons comment cette méthode est liée à des algorithmes modernes, et nous l'utilisons pour proposer les prémices d'une théorie exacte de la factorisation de matrices à rang extensif. Pour cela nous exploitons la connection forte qui relie la physique des systèmes désordonnés et les statistiques en grande dimension, un sujet de recherche qui suscite un intérêt croissant depuis les années 1990. Dans une seconde partie nous poussons cette correspondance plus loin et utilisons des outils heuristiques de physique théorique, comme la méthode des répliques, associés à des outils probabilistes et des algorithmes de passage de message, pour décrire les limites fondamentales d'une grande catégorie de problèmes d'apprentissage. Nous appliquons cette analyse à des réseaux de neurones, à l'extraction de phase, ainsi que pour étudier l'influence de la structure des données sur les procédures d'inférence. Enfin nous proposons une direction alternative, une approche topologique au problème d'inférence en haute dimension: en utilisant des outils de géométrie différentielle stochastique et de matrices aléatoires, nous prouvons des formules exactes décrivant la structure des paysages d'énergie optimisés par les algorithmes d'apprentissage.

MOTS CLÉS

Physique statistique, Apprentissage automatique, Statistiques en haute dimension, Théorie de l'information, Théorie des matrices aléatoires, Optimisation non convexe.

ABSTRACT

The past decade saw an intensification of the deluge of data available to learning algorithms, which allowed for the development of modern artificial intelligence techniques. These methods rely on the optimization of a very large number of internal parameters using gigantic amounts of data, and now provide state-of-the-art algorithms for tasks as diverse as image classification, natural language processing, or speech recognition, and regularly achieve super-human performances. This exacerbated research efforts to build a mathematically sound theory of data science able to explain the extraordinary efficiency of these procedures, and has led to a surge of interest for high-dimensional statistics (i.e. when the amount of data and the number of parameters are both very large). In this dissertation we analyze a few pieces of this immense puzzle through the prism of statistical physics, borrowing also often from probability and random matrix theory, and we propose three approaches to the high-dimensional learning problem. In the first one we revisit high-temperature expansions, an archetypal method of statistical physics. We show how this classical approach is related to modern algorithms, and use it to pave the way towards an exact theory of extensive-rank matrix factorization. Our theory leverages the intimate relation between the statistical physics of disordered systems and high-dimensional statistics, a connection which has been a growing subject of research since the 1990s. Our second approach pushes further this correspondence as we leverage heuristic tools of theoretical physics such as the replica method, along with modern probabilistic methods and message-passing algorithms, to describe the fundamental limits of a wide class of high-dimensional learning problems. We apply our analysis to neural networks, phase retrieval, and to study the influence of data structure on the optimal learning procedures. In a third part we take an alternative route and consider a topological approach to the problem of learning in high dimension. Using tools of random differential geometry and random matrix theory we prove exact formulas describing the structure of the high-dimensional landscapes optimized by learning algorithms.

KEYWORDS

Statistical physics, Machine learning, High-dimensional statistics, Information theory, Random matrix theory, Non-convex optimization.