Fitting ellipsoids to random points

Antoine Maillard

ETH zürich

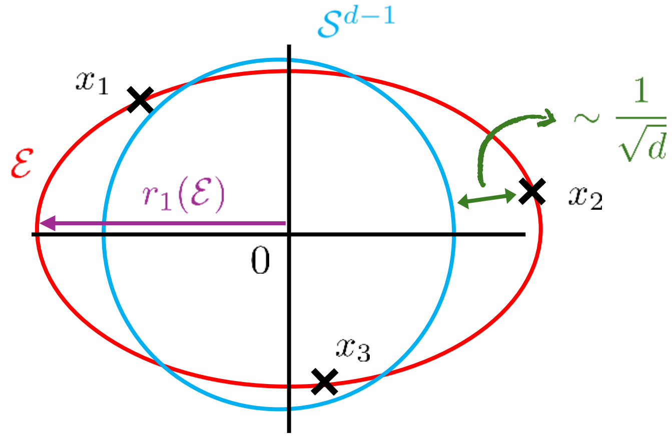
- arXiv:2307.01181 *(w. A. Bandeira, D. Kunisky, S. Mendelson & E. Paquette)*
- arXiv:2310.01169 *(w. D. Kunisky)* – IEEE Trans. Inf. Theory '24
- arXiv:2310.05787 *(w. A. Bandeira)*

WDI^2 Innsbruck – June 28th 2024

Fitting ellipsoids to random points

$$x_1, \cdots, x_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d/d)$$
$$n, d \to \infty$$

Does $\mathcal{E}$ exist ?



Ellipsoid Fitting Property

$$\mathbb{P}[\exists S \in \mathbb{R}^{d\times d} : S \succeq 0 \text{ and } x_i^\top S x_i = 1 \text{ for all } i \in [n]] \; ?$$

Principal axes of $\mathcal{E}$ $\Longleftrightarrow$ Eigenspaces of S

$$r_i(\mathcal{E}) = \lambda_i(S)^{-1/2}$$

- EFP is a **semidefinite program**
- Norm fluctuations are critical
  - $x_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\{\pm 1/\sqrt{d}\}^d)$
  - $x_i \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\mathcal{S}^{d-1})$

  } $S = \mathrm{I}_d$ is always a solution

Fitting ellipsoids to random points

Ellipsoid Fitting Property

$$p(n,d) := \mathbb{P}[\exists S \in \mathbb{R}^{d\times d} : S \succeq 0 \text{ and } x_i^\top S x_i = 1 \text{ for all } i \in [n]]$$

Some motivations

Potechin & al '22

- Low-rank matrix decomposition — Saunderson & al '12 ; '13 ; '13

  Recommendation systems, community detection, ...

  $X = D^\star + L^\star \in \mathbb{R}^{n\times n}$

  Diagonal ($D^\star$); $\succeq 0$ + low-rank ($L^\star$)

  $$\mathrm{MTFA} := \min_{\substack{D,L\,:\,X=D+L \\ L\succeq 0}} \mathrm{Tr}(L)$$

  $\mathrm{col}(L^\star) \sim \mathrm{Unif}[r-\text{dim subspaces}] \Rightarrow \mathbb{P}[\text{MTFA recovers } (L^\star, D^\star)] = p(n, n-r)$

- Independent Components Analysis — Podosinnikova & al '19

  Signal processing

- Discrepancy of random matrices — Potechin & al '22

  SDP lower bounds certification

- Neural networks with quadratic activations — **M.**, Troiani, Martin, Krzakala, Zdeborová '24

  Optimization, machine learning,...

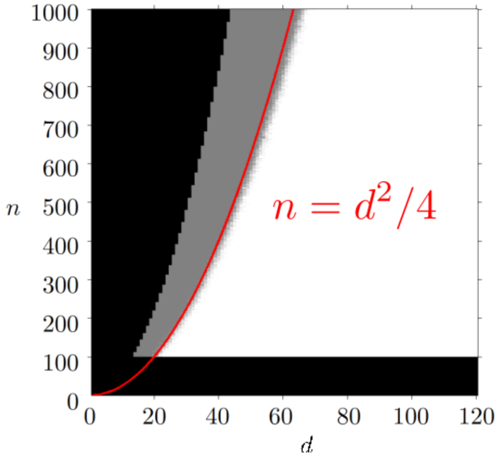
The ellipsoid fitting conjecture

Ellipsoid fitting is a **semidefinite program**

⇩

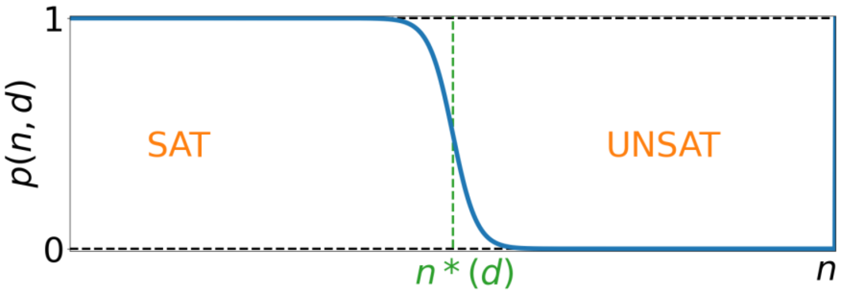
Convex problem + efficient solvers

■ : No simulation ■ : No solutions □ : Solutions exist



Saunderson, James, et al. *SIAM Journal on Matrix Analysis and Applications* 2012

$$p(n,d) = \mathbb{P}[\text{An ellipsoid fit exists}]$$

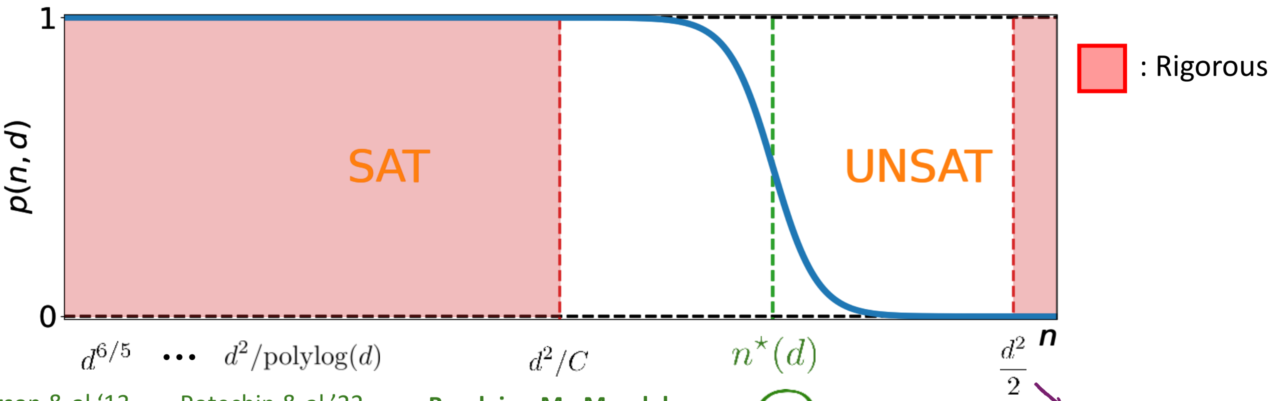


Open conjecture

$$\lim_{d\to\infty} \frac{n^\star(d)}{d^2} = \frac{1}{4}$$

The ellipsoid fitting conjecture: what is known

Conjecture $\lim_{d\to\infty} \frac{n^\star(d)}{d^2} = \frac{1}{4}$



Progress on **lower bounds**

Saunderson & al '13

Potechin & al '22
Kane & al '22

Bandeira, M., Mendelson & Paquette ' 23 ; Hsieh & al '23 ; Tulsiani & Wu '23

?

Dimension counting
$\dim(\{S = S^\top\}) \simeq d^2/2$

I   II

This talk

Lower bounds [Bandeira, **M.** , Mendelson & Paquette '23]

Goal: $\lim_{d\to\infty} \mathbb{P}[\text{Ellipsoid fit exists}] = 1$ for $n < n_c(d)$

w.h.p. = "with high probability" $\Leftrightarrow$ $\mathbb{P}[\cdot] = 1 - o_d(1)$

Existing works on EFP rely on an **explicit estimator**:

- $\hat{S}_{\mathrm{LS}} := \underset{\{x_i^\intercal S x_i = 1\}}{\arg\min} \|S\|_F$

  [Potechin & al '22]

  Theorem: $\hat{S}_{\mathrm{LS}} \succeq 0$ w.h.p. if $n \lesssim d^2/\mathrm{polylog}(d)$

  Non-rigorous analysis shows this holds for $n \leq d^2/10$

  [**M.** & Kunisky '23]

- $\hat{S}_{\mathrm{IP}} := \mathrm{I}_d + \sum_{i=1}^{n} q_i x_i x_i^\intercal$

  $\{x_i^\intercal \hat{S}_{\mathrm{IP}} x_i = 1\}_{i=1}^n$ n linear equations in $q \in \mathbb{R}^n$

Theorem: $\hat{S}_{\mathrm{IP}} \succeq 0$ (w.h.p.) if

- $n \lesssim d^2/\mathrm{polylog}(d)$ [Kane & Diakonikolas '22]
- $n \leq d^2/C$ [Bandeira, **M.**, Mendelson & Paquette '23]

Numerically $C \simeq 10$

Similar result obtained w. different estimates in [Hsieh&al'23 ; Tulsiani & Wu '23]

Lower bounds – Sketch of proof [Bandeira, **M.** , Mendelson & Paquette '23]

$\omega_i \sim \mathrm{Unif}(\mathcal{S}^{d-1})$

Rotation invariance ⇒ $x_i = \sqrt{\tau_i}\omega_i \sim \mathcal{N}(0, \mathrm{I}_d/d)$

Define $\begin{cases} \Theta_{ij} := \langle \omega_i, \omega_j \rangle^2 \\ \Gamma := \mathrm{Diag}(\{\tau_i\}_{i=1}^n) \end{cases}$

$\hat{S}_{\mathrm{IP}} := \mathrm{I}_d + \sum_{i=1}^n q_i x_i x_i^\intercal$ + $\{x_i^\intercal \hat{S}_{\mathrm{IP}} x_i = 1\}_{i=1}^n$ ⇒ $q = \Gamma^{-1}\Theta^{-1}(\Gamma^{-1}\mathbf{1}_n - \mathbf{1}_n)$

We show $\|\cdot\|_{\mathrm{op}} \leq 1$

Goal: $\left\| \sum_{i=1}^n [\Theta^{-1}(\Gamma^{-1}\mathbf{1}_n - \mathbf{1}_n)]_i \omega_i \omega_i^\intercal \right\|_{\mathrm{op}.} \leq 1$

i.i.d. independent of ω_i

- Key difficulty: controlling $\|\Theta^{-1}\|_{\mathrm{op}}$ ⚠
- Rest of the proof: classical ε-net argument.

$\Theta_{ij} = \langle \omega_i \omega_i^\intercal, \omega_j \omega_j^\intercal \rangle$

Gram matrix of sub-exp. random vectors in $\mathbb{R}^p$

$p := \binom{d+1}{2}$

⇒ **Key lemma**

$\|\Theta - \mathbb{E}\Theta\|_{\mathrm{op}} \lesssim \sqrt{\frac{n}{d^2}}$ w.h.p.

[Bartl & Mendelson '22]

⇒ $\|\Theta^{-1}\|_{\mathrm{op}} \leq 2$ w.h.p. for small enough $\frac{n}{d^2}$ ■

Statistical physics tools for ellipsoid fitting [M. & Kunisky '23]

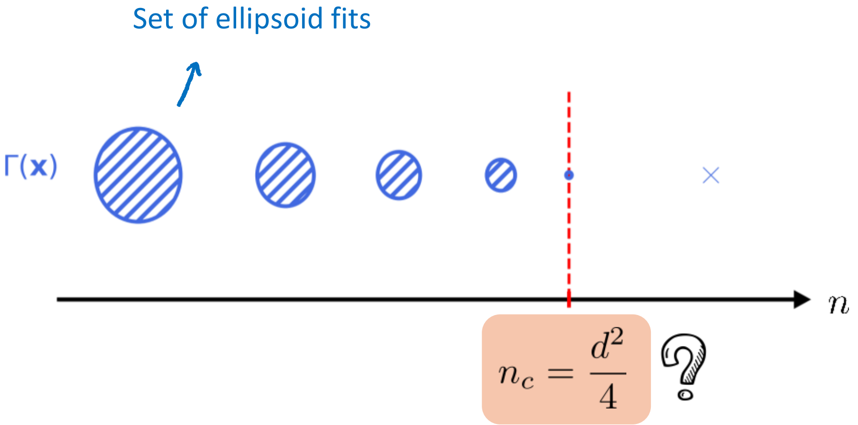
Ellipsoid Fitting Property

$$\mathbb{P}[\exists S \in \mathbb{R}^{d\times d} \;:\; S \succeq 0 \text{ and } x_i^\top S x_i = 1 \text{ for all } i \in [n]] \;?$$

We see EFP as a **Random Constraint Satisfaction Problem**

$$\begin{cases} S \succeq 0 & \to \text{"spectral" constraint} \\ \{x_i^\top S x_i = 1\}_{i=1}^n & \end{cases}$$

x_i → **"disordered" model**



Volume of solutions / **"Partition function"**

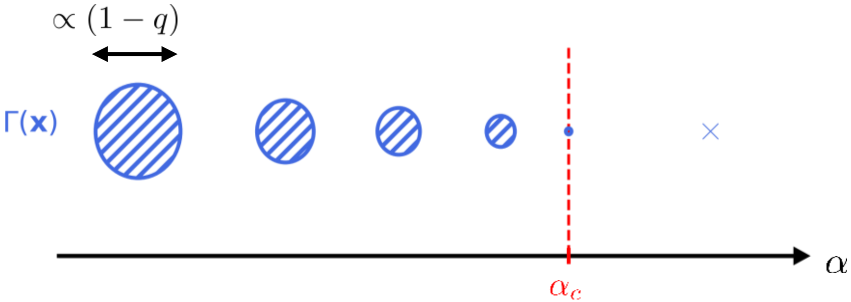
$\mathrm{supp}(P_0) \subseteq \mathcal{S}_d^+$

$$\mathcal{Z} := \int P_0(\mathrm{d}S) \prod_{i=1}^n \delta(x_i^\top S x_i - 1)$$

Statistical physics tools for ellipsoid fitting [M. & Kunisky '23]

$$n/d^2 \to \alpha > 0$$

$$\mathcal{Z} := \int P_0(\mathrm{d}S) \prod_{i=1}^{n} \delta(x_i^\top S x_i - 1)$$



Replica method

$$\frac{1}{d^2}\mathbb{E}\log\mathcal{Z} \to \sup_{q\in[0,1]} \sup_{\mu\in\mathcal{M}_1^+(\mathbb{R})} \left[F(\alpha,q,\mu) + I_{\mathrm{HCIZ}}\left(\frac{1}{\sqrt{1-q}},\mu,\sigma_{\mathrm{s.c.}}\right)\right]$$

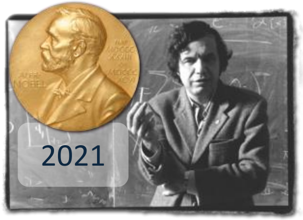
"Overlap"

Typical spectrum of solutions (ellipsoid shape)

$$I_{\mathrm{HCIZ}}(\theta,A,B) := \lim_{d\to\infty}\frac{1}{d^2}\log\int_{\mathcal{O}(d)} \mathcal{D}O \exp\{\theta \mathrm{Tr}[OAO^\top B]\}$$

Hard asymptotic expressions via PDEs [Matytsin '94 ; Guionnet&al'02]

Non-rigorous analytical method from statistical physics



Giorgio Parisi

$\alpha \to \alpha_c$
$q \to 1$

+

"Dilute" expansion ($\theta \to \infty$) of $I_{\mathrm{HCIZ}}(\theta,A,B)$ [Bun & al '16]

→

$$\alpha_c = \frac{1}{4}$$

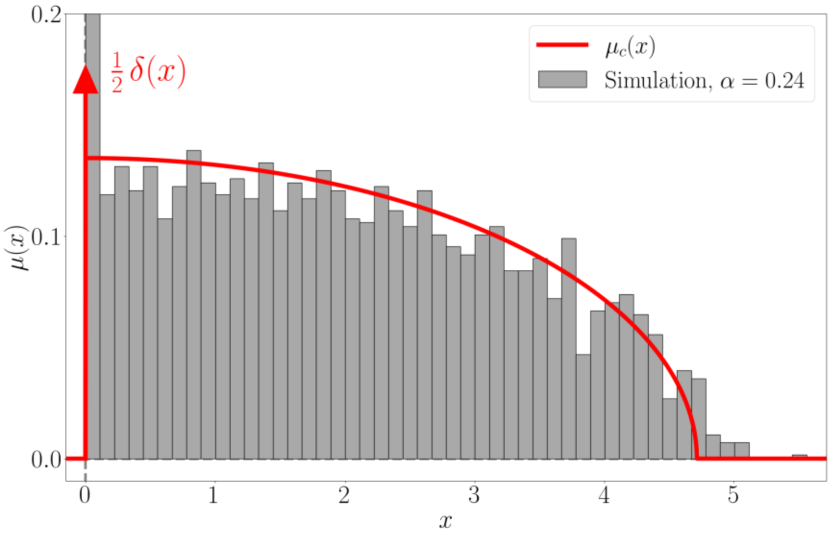
+

- Computation of typical μ
- Extensions to non-Gaussian x_i
- …

Statistical physics tools for ellipsoid fitting [M. & Kunisky '23]

Spectrum of solutions / Shape of ellipsoids

Near the transition $\alpha \uparrow 1/4$



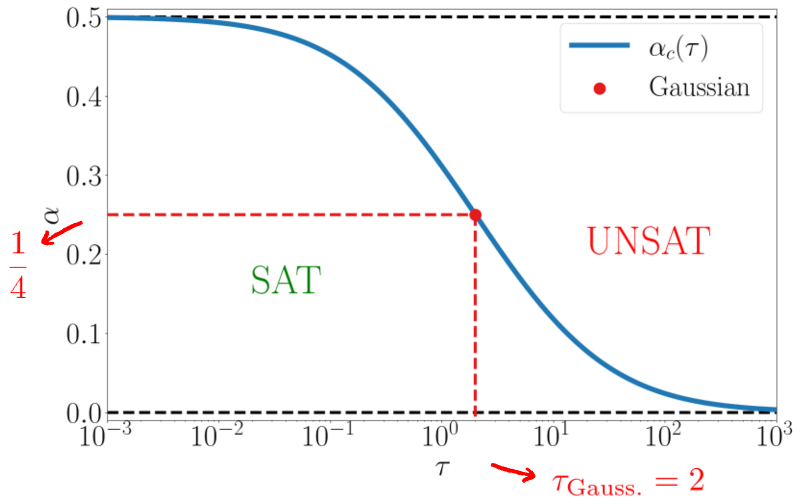
- <u>Truncated semicircular distribution</u>
- As $\alpha \uparrow 1/4$, ellipsoid fits are "cylinders" in $d/2$ directions !

Generalization to non-Gaussian random vectors

$$x_i = \sqrt{r_i}\omega_i$$

$\omega_i \sim \mathrm{Unif}(\mathcal{S}^{d-1})$

$\mathbb{E}[r_i] = 1$ + $\mathrm{Var}(r_i) = \frac{\tau}{d}$



Larger norm fluctuations ⇒ Ellipsoid fits harder to find

Mathematical physics for ellipsoid fitting [M. & Bandeira '23]

Two-steps proof

I: • "**Gaussian universality**" lemma : $\frac{1}{n}\log \mathcal{Z} \simeq \frac{1}{n}\log \mathcal{Z}_G$

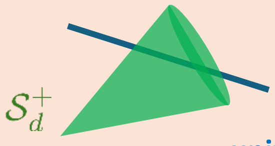
[Goldt & al '22, Montanari & Saeed '22, Hu & Lu '22, …]

$x_i^\top S x_i \longrightarrow \mathrm{Tr}(SG_i)$ ← Gaussian matrix

$$\mathcal{Z} := \int P_0(\mathrm{d}S)\prod_{i=1}^n \delta(x_i^\top S x_i - 1) \longrightarrow \mathcal{Z}_G := \int P_0(\mathrm{d}S)\prod_{i=1}^n \delta(\mathrm{Tr}(SG_i) - 1)$$

II: • **Random convex geometry** tools for $\mathcal{Z}_G$

Extensions of Gordon's **min-max theorem**
[Gordon '88, Amelunxen & al'14]



Theorem: The problem associated to $\mathcal{Z}_G$ is
- SAT (whp) if $n \leq (1-\varepsilon)\omega(\mathcal{S}_d^+)^2$
- UNSAT (whp) if $n \geq (1+\varepsilon)\omega(\mathcal{S}_d^+)^2$

Gaussian width

$$\omega(\mathcal{S}_d^+) := \mathbb{E} \max_{\substack{S \succeq 0 \\ \|S\|_F = 1}} \mathrm{Tr}[GS]$$

$$\omega(\mathcal{S}_d^+) \sim_{d\to\infty} \frac{d}{2} \Rightarrow n^\star(\mathcal{Z}_G) \sim \frac{d^2}{4}$$

Mathematical physics for ellipsoid fitting [M. & Bandeira '23]

II

I: "**Gaussian universality**" lemma + II: **Random convex geometry** tools

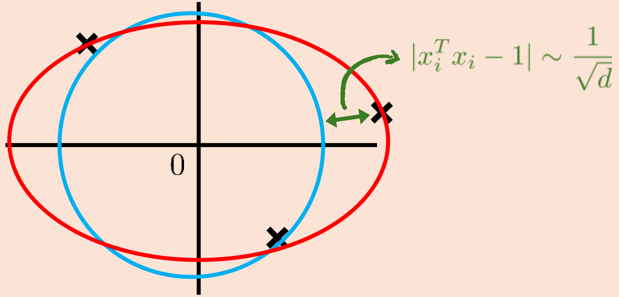
Theorem

EFP$_{\varepsilon,\, M}$ **:** $\exists S \in \mathbb{R}^{d\times d} \ : \ \mathrm{Sp}(S) \subseteq [0, M] \text{ and } \frac{1}{n}\sum_{i=1}^{n} |x_i^\top S x_i - 1| \leq \frac{\varepsilon}{\sqrt{d}}$

EFP $=$ **EFP**$_{0,\, \infty}$

$^n/d^2 \to \alpha$

- $\alpha < {}^1/4 \quad \exists M_\alpha \ : \ \forall \varepsilon > 0, \ \mathbb{P}[\mathbf{EFP}_{\varepsilon, M_\alpha}] \to_{d\to\infty} 1$
- $\alpha > {}^1/4 \quad \exists \varepsilon_\alpha \ : \ \forall M > 0, \ \mathbb{P}[\mathbf{EFP}_{\varepsilon_\alpha, M}] \to_{d\to\infty} 0$



Ellipsoid fitting: summary

1. Best-known **lower bound** $n^\star(d) \geq \frac{d^2}{C}$ — Bandeira, **M.**, Mendelson & Paquette ’23

2. Refinement and extension of the conjecture to **non-Gaussian points.** — **M.** & Kunisky ’23
   to appear in IEEE Trans. Inf. Theory

3. Theorem: $n^\star(d) = \frac{d^2}{4}$ in **approximate ellipsoid fitting**. — **M.** & Bandeira ’23
   First rigorous characterization of the transition

- Strengthen proof to **exact** ellipsoid fitting ?
- Extension to **other high-dimensional semidefinite programs** ?
- Relevance of proof techniques for **learning in neural networks.**
  (joint work with E. Troiani, S. Martin, F. Krzakala, L. Zdeborová)

THANK YOU !