Extensive Rank Matrix Estimation

▷ **Matrix factorization—** from a noisy measurement of the product of two matrices the aim is to estimate the original matrices, that is, the components that formed the product

$$\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{X} + \boldsymbol{R}\,,$$

with $\boldsymbol{F} \in \mathbb{R}^{N\times R}$, $\boldsymbol{X} \in \mathbb{R}^{R\times M}$ and $R = \Theta(N) = \Theta(M)$.

Examples: dictionary learning, blind source separation, matrix completion, PCA, ...

▷ **A simpler problem**— the Bayes-optimal *symmetric* version with a Gaussian prior on $\boldsymbol{X}$ under Gaussian noise $\boldsymbol{W}$ of variance Δ:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{X}^\intercal + \sqrt{\Delta}\boldsymbol{W}\,, \qquad \boldsymbol{X} \in \mathbb{R}^{N\times R}\,, \qquad \alpha := \frac{N}{R} \xrightarrow{N\to\infty} \Theta(1)$$

Note: $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\intercal$ is Wishart distributed

$$P(\boldsymbol{S}) \propto \det(\boldsymbol{S})^{(R-N-1)/2} \mathrm{e}^{-\mathrm{Tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}/2}$$

$\Rightarrow$ higher order correlations among the elements of $\boldsymbol{S}$.

▷ **Motivation**— accessible inference problems typically either decouple in some fashion or can be reduced to a single scalar order parameter.

- For extensive spin models, this is no longer the case ⇒ overlap does not concentrate.

Question: how to analyze spin models with $\Theta(N)$ elements per spin?

Idea: combine replicas with random matrix theory.

Replica Approach— A Random Matrix Problem in the Replica Space

▷ **Posterior probability**— for $\boldsymbol{X}$, given $\boldsymbol{Y}$

$$P(\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{1}{Z(\boldsymbol{Y})} \prod_{\mu \le N} \prod_{k \le R} P(x_{\mu k}) \prod_{1 \le \mu \le \nu \le N} P(y_{\mu\nu} \mid \sum_k x_{\mu k} x_{\nu k}) \, .$$

with $P(x_{\mu k})$ and $P(y_{\mu\nu} \mid \sum_k x_{\mu k} x_{\nu k})$ both Gaussians.

▷ **Aim**— compute $\mathbb{E}_{\boldsymbol{Y}} \ln Z(\boldsymbol{Y})$, with

$$Z(\boldsymbol{Y}) = c \cdot \int \mathrm{d}\boldsymbol{X} \exp\left(-\frac{N}{2} \mathrm{Tr} \boldsymbol{X}\boldsymbol{X}^\intercal\right) \exp\left[-\frac{N}{4\Delta} \mathrm{Tr}\left(\boldsymbol{Y} - \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\intercal\right)^2\right] .$$

Problem: disorder average over logarithm and quartic terms.

▷ **Replicas**— resolve disorder average by *replica trick* and eliminate quartic term by *Hubbard-Stratonovich transform*.

▷ One obtains matrices $\boldsymbol{Q}^{ab} = (Q^{ab}_{kk'})_{1\leq k,k'\leq R}$ instead of scalars.

$$\mathbb{E}_{\boldsymbol{Y}}\left[\tilde{Z}(\boldsymbol{Y})^n\right] \propto \int \mathrm{d}\underline{\underline{\boldsymbol{Q}}}\, \mathrm{e}^{-N\frac{\Delta}{4}\mathrm{Tr}\underline{\underline{\boldsymbol{Q}}}^2} \prod_{\mu} \int \mathrm{d}\underline{\boldsymbol{x}}_\mu\, \mathrm{e}^{-\frac{1}{2}\underline{\boldsymbol{x}}_\mu^\intercal\left(\underline{\underline{\mathbf{1}}}-\underline{\underline{\boldsymbol{Q}}}\right)\underline{\boldsymbol{x}}_\mu}$$

- Integration factorizes over the rows, *not* w.r.t. both indices.

▷ Denote

$$\underline{\underline{\boldsymbol{Q}}} = \begin{pmatrix} \mathbf{0} & \boldsymbol{Q}^{01} & \cdots & \boldsymbol{Q}^{0n} \\ \boldsymbol{Q}^{10} & \mathbf{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ \boldsymbol{Q}^{n0} & \cdots & & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(n+1)R\times(n+1)R}, \qquad \underline{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{x}^0 \\ \boldsymbol{x}^1 \\ \vdots \\ \boldsymbol{x}^n \end{pmatrix} \in \mathbb{R}^{(n+1)R}.$$

▷ **Overlap between two samples—**

$$m^{ab} := \sum_k x^a_{\mu k} x^b_{\mu k} \simeq \mathrm{Tr}\mathbb{E}_{\underline{\underline{\boldsymbol{Q}}}}\left[\boldsymbol{x}^a_\mu (\boldsymbol{x}^b_\mu)^\mathsf{T}\right]$$

▷ **Auxiliary Problem—** to determine distribution of $\underline{\underline{\boldsymbol{Q}}}$

$$\mathbb{E}_{\boldsymbol{Y}}\left[\tilde{Z}(\boldsymbol{Y})^n\right] \propto \int \mathrm{d}\underline{\underline{\boldsymbol{Q}}} \exp\left(-N\frac{\Delta}{4}\mathrm{Tr}\underline{\underline{\boldsymbol{Q}}}^2 - \frac{N}{2}\ln\det\left(\underline{\underline{\mathbf{1}}} - \underline{\underline{\boldsymbol{Q}}}\right)\right)$$

▷ Expect huge degeneracy, due to rotational symmetry.

$$\boldsymbol{Q}^{ab} = \boldsymbol{O}^{ab}\boldsymbol{L}^{ab}(\boldsymbol{O}^{ab})^\mathsf{T} \quad \Rightarrow \quad \mathrm{d}\boldsymbol{Q}^{ab} \propto (\mathrm{d}\boldsymbol{O}^{ab}) \prod_{i<j}\left|l^{ab}_i - l^{ab}_j\right| (\mathrm{d}\boldsymbol{L}^{ab}) \, .$$

▷ **Matrix action**— in terms of the eigenvalues $\boldsymbol{L}^{ab}$

$$\mathbb{E}_{\boldsymbol{Y}}\left[\tilde{Z}(\boldsymbol{Y})^n\right] \propto \int \mathrm{d}\underline{\underline{\boldsymbol{L}}}\, \mathrm{e}^{-N^2 S[\underline{\underline{\boldsymbol{L}}}]}$$

with

$$S[\underline{\underline{\boldsymbol{L}}}] = \frac{\Delta}{4}\frac{1}{N}\sum_{a\neq b}\mathrm{Tr}(\boldsymbol{L}^{ab})^2 - \frac{1}{N^2}\sum_{a\neq b}\sum_{i<j}\ln|l_i^{ab} - l_j^{ab}| - G[\underline{\underline{\boldsymbol{L}}}]$$

and

$$G[\underline{\underline{\boldsymbol{L}}}] = \frac{1}{N^2}\ln\int \mathrm{D}\underline{\underline{\boldsymbol{O}}}\exp\Big(-\frac{N}{2}\ln\det(\underline{\underline{\mathbf{1}}} - \underline{\underline{\boldsymbol{Q}}})\Big)$$

▷ **Saddlepoint approximation**— variation w.r.t. the eigenvalues l_i^{ab}

$$\frac{\Delta}{2} l_k^{ab} - \frac{1}{N}\sum_{\substack{1\leq i\leq R\\ i\neq k}}\frac{1}{l_k^{ab} - l_i^{ab}} - N\frac{\partial G}{\partial l_k^{ab}} = 0\,, \qquad \forall k, \forall a\neq b$$

Replica Symmetric (RS) Solution

▷ **RS Ansatz**— same overlap between the replicas $a \neq b$

$$\underline{\underline{\boldsymbol{Q}}} = \begin{pmatrix} \mathbf{0} & \boldsymbol{OLO}^{\mathsf{T}} & \cdots & \boldsymbol{OLO}^{\mathsf{T}} \\ \boldsymbol{OLO}^{\mathsf{T}} & \mathbf{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ \boldsymbol{OLO}^{\mathsf{T}} & \cdots & & \mathbf{0} \end{pmatrix} .$$

▷ $G[\underline{\underline{\boldsymbol{L}}}]$ becomes invariant under the orthogonal transformation.

▷ Saddlepoint approximation leads to equation on the eigenvalues only.

▷ **Continuous limit**— integration over $\boldsymbol{L} \to$ integration over $\rho(\lambda)$.

$$\mathbb{E}_{\boldsymbol{Y}}\left[\tilde{Z}(\boldsymbol{Y})^n\right] \overset{\mathrm{RS}}{\propto} \int \mathcal{D}\rho(\lambda)\, \mathrm{e}^{-N^2\, n\, \mathcal{F}_{\mathrm{RS}}[\rho(\lambda)]}$$

with free energy

$$\mathcal{F}_{\mathrm{RS}}[\rho(\lambda)] = \frac{\Delta}{4\alpha}\int \mathrm{d}\lambda\rho(\lambda)\,\lambda^2 - \frac{1}{2\alpha}\int \mathrm{d}\lambda\rho(\lambda)\,\lambda + \frac{1}{2\alpha}\int \mathrm{d}\lambda\rho(\lambda)\ln(1+\lambda) - \frac{1}{2\alpha^2}\int\int \mathrm{d}\lambda\mathrm{d}\lambda'\rho(\lambda)\rho(\lambda')\ln|\lambda-\lambda'|\,.$$

▷ **RS saddle point**— leads to singular integral equation

$$\boxed{\underbrace{\frac{1}{2}\left(\Delta\lambda - \frac{\lambda}{\lambda+1}\right)}_{\text{eff. potential force}} = \underbrace{\frac{1}{\alpha}\fint \mathrm{d}\lambda'\,\frac{\rho(\lambda')}{\lambda-\lambda'}}_{\text{repulsive force}}\,.}$$

Solving the RS saddle point equations

▷ **Random matrix theory**— RS saddle point ⇔ RMT saddle point

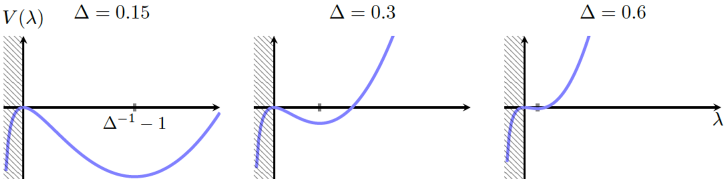
$$V'(\lambda) = \frac{1}{\alpha} \fint \mathrm{d}\lambda' \frac{\rho(\lambda')}{\lambda - \lambda'} .$$

<u>Solution</u>: Riemann-Hilbert Ansatz [2] or Tricomi's theorem [3].

▷ x-**MMSE**:

$$x\text{-MMSE} = 2 - 2 \int \mathrm{d}\lambda \rho(\lambda) \frac{\lambda}{1+\lambda} .$$

▷ Enforce $\lambda \geq 0$.



Phase Diagram

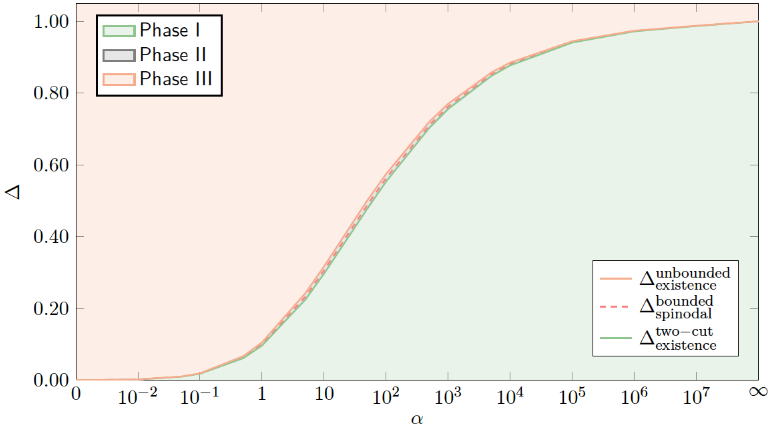
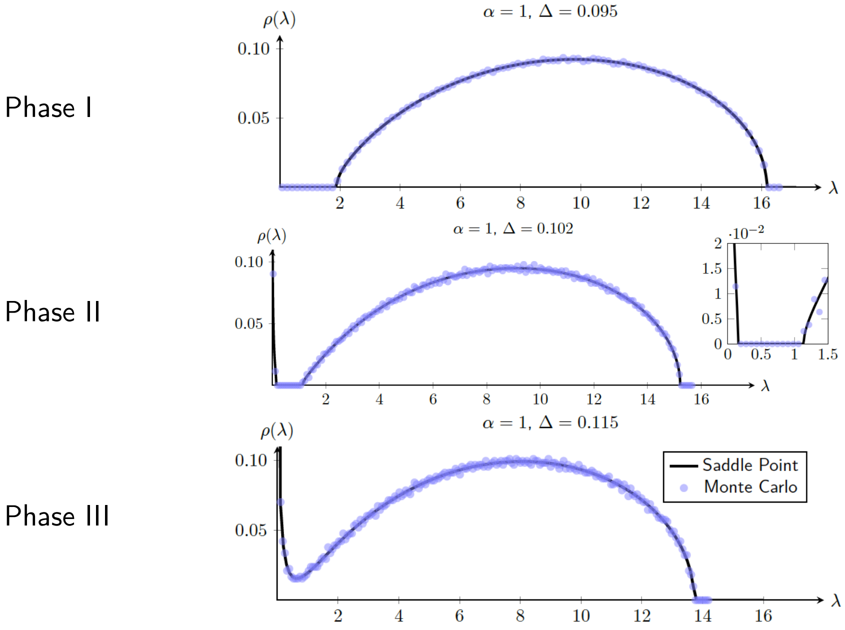


Figure: In Phase I the solution is supported along a single cut and is bounded on both ends of it. In Phase II it has support on two separate cuts. In Phase III the solution single compact support and is unbounded at zero.



Phase I
Phase II
Phase III
α = 1, Δ = 0.095
α = 1, Δ = 0.102
α = 1, Δ = 0.115
ρ(λ)
λ
·10^−2
Saddle Point
Monte Carlo

References

[1] Kabashima, Y., Krzakala, F., Mézard, M., Sakata, A., & Zdeborová, L. (2016). *Phase Transitions and Sample Complexity in Bayes-Optimal Matrix Factorization*. IEEE Transactions on Information Theory 62(7), 4228–4265.

[2] Brézin, E., Itzykson, C., Parisi, G., & Zuber, J. B. (1978). *Planar diagrams*. Communications in Mathematical Physics 59, 35–51.

[3] Tricomi, F. G. (1957). *Integral equations*. Interscience Publishers.

[4] Gakhov, F. D. (1966). *Boundary value problems*. Pergamon Press.

Theoretical Limits in Extensive Rank Matrix Estimation

Christian Schmidt, Pierfrancesco Urbani and Lenka Zdeborová
Institut de Physique Théorique, CEA and Université Paris Saclay