Inference from heterogeneous pairwise data

Galen Reeves

Department of ECE and Department of Statistical Science
Duke University

August 2023

Pairwise inference model

- Latent variables $(x_1, \ldots, x_N) \in \mathcal{X}^N$.
- Observations through memoryless channel P with input alphabet $\mathcal{X} \times \mathcal{X}$

$$Y_{ij} \overset{\text{ind}}{\sim} P(\cdot \mid x_i, x_j), \quad i, j = 1, \ldots N$$

Spin glasses

Sherrington-Kirkpatrick model (with planted signal)

$$Y_{ij} \sim \mathsf{N}\Big(\sqrt{\tfrac{\lambda}{N}} x_i x_j, 1\Big), \quad \lambda \geq 0, \quad x_i \in \{\pm 1\}$$

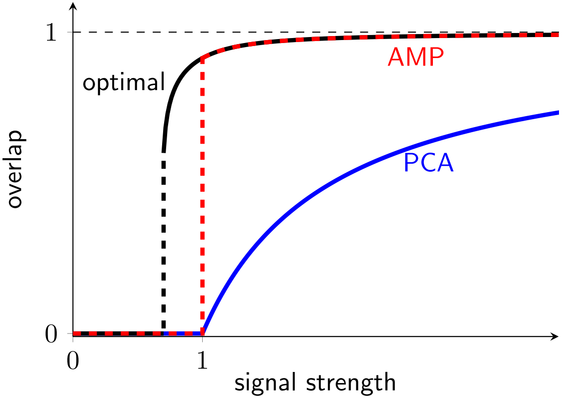
Stochastic block model

$$Y_{ij} \sim \mathsf{Bern}(Q_{x_i, x_j}), \quad Q \in [0,1]^{K \times K}, \quad x_i \in \{1, \ldots, K\}$$

Spiked covariance model

$$Y_{ij} \sim \mathsf{N}\Big(\sqrt{\tfrac{\lambda}{N}} \langle u_i, v_j \rangle, 1\Big), \quad \lambda \geq 0, \qquad u_i, v_j \in \mathbb{R}^r$$

Statistical & computational limits



Why mutual information?

$$I(\boldsymbol{X};\boldsymbol{Y}) = D(P_{XY} \,\|\, P_X \otimes P_Y) = \int dP_{XY} \log \frac{dP_{XY}}{d(P_X \otimes P_Y)}$$

I-MMSE relationship: If $\boldsymbol{X} \in \mathbb{R}^{N \times d}$ then

$$\frac{d}{dt} I\big(\boldsymbol{X};\boldsymbol{Y}, \sqrt{t}\boldsymbol{X} + \boldsymbol{Z}\big)\Big|_{t=0} = \mathbb{E}\big[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \boldsymbol{Y}]\|_F^2\big]$$

where $\boldsymbol{Z} \in \mathbb{R}^{N \times d} \perp\!\!\!\perp (\boldsymbol{X}, \boldsymbol{Y})$ has IID standard Gaussian entries.

Hence, formula for asymptotic mutual information (with side information) gives formula for asymptotic minimum mean square error almost everywhere*.

* Limit is not differentiable at phase transitions

Channel universality

Classical scaling (fixed p large n) Local asymptotic normality & Fisher information

$$\log \frac{dP_{n,\theta+\epsilon_n u}}{dP_{n,\theta}} \xrightarrow[n\to\infty]{d} \log \frac{dQ_u}{dQ_0}, \qquad \underbrace{Q_u = \mathsf{N}(I_\theta u, I_\theta)}_{\text{Gaussian model}}$$

High-dimensional scaling Korada-Montanari 2011, Deshpande-Abbe-Montanari 2015
Krzakala-Xu-Zdeborová 2016, Lesieur-Krzakala-Zdeborová 2017, **R.**-Mayya-Volfovsky 2019
Guionnet-Ko-Krzakala-Zdeborová 2023

Compare free energy / mutual information for prior distribution π on θ

$$\left| \mathbb{E}_{P_\theta} \log \int \frac{dP_{n,\theta+r_n u}}{dP_{n,\theta}} \pi(du) - \mathbb{E}_{Q_0} \log \int \frac{dQ_u}{dQ_0} \pi(du) \right| = o(n)$$

- Gaussian comparison holds for dense SBM but not for sparse SBM

Model for Today's talk

- Assume $\mathcal{X}$ is (or can be embedded into) a subset of $\mathbb{R}^d$. (d is fixed)
- $\boldsymbol{X} = (x_1, \ldots, x_N)^\top$ is $N \times d$ matrix with IID rows & finite fourth moments.

Family of channels $(P_N)_{N\in\mathbb{N}}$. Approximate by linear Gaussian model

$$P_N(\cdot \mid x_i, x_j) \quad \approx \quad \mathsf{N}\Big(\sqrt{\tfrac{1}{N}}\, B^\top (x_i \otimes x_j), \mathrm{I}_L\Big), \quad B \in \mathbb{R}^{d^2 \times L}$$

- Includes bounded degree polynomials in Gaussian noise via lifting

$$f(\theta_i, \theta_j) = b^\top (x_i \otimes x_j), \qquad x_i = (1, \theta_i, \theta_i^2, \ldots \theta_i^{d-1})^\top$$

- Includes groupwise heteroskedasticity via augmentation

$$x_i = \underbrace{(0, \ldots, \theta_i, \ldots, 0)^\top}_{\text{position indexed by group}}$$

Table of Contents

Entropy / free energy with IID prior

- Rows of $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ are IID copies of $X_0 \in \mathbb{R}^d$ with finite fourth moments.

$$D(\boldsymbol{Y} \,\|\, \boldsymbol{Z}) = \underbrace{\mathbb{E}\left[\log \int e^{\langle \boldsymbol{x}^{\otimes 2} B, \boldsymbol{Y}\rangle - \frac{1}{2}\|\boldsymbol{x}^{\otimes 2} B\|_F^2} P_0^{\otimes N}(d\boldsymbol{x})\right]}_{\text{original problem is } N \times d}, \quad \boldsymbol{Y} = \boldsymbol{X}^{\otimes 2} B + \boldsymbol{Z}$$

- Entropy function $\mathcal{D} \colon \mathbb{S}_+^d \to [0, \infty)$ defined by

$$\begin{aligned}\mathcal{D}(R) &\coloneqq D\big(R^{1/2} X_0 + Z_0 \,\|\, Z_0\big), \\ &= \underbrace{\mathbb{E}\left[\log \int e^{\langle x, R X_0 + Z_0\rangle - \frac{1}{2} x^\top R x} P_0(dx)\right]}_{\text{easy problem is } d \times 1}, \quad X_0 \perp\!\!\!\perp Z_0 \sim \mathsf{N}(0, I_d)\end{aligned}$$

Symmetric spiked matrix model

$$\boldsymbol{Y} = \sqrt{\frac{\lambda}{N}}\,\boldsymbol{X}\boldsymbol{X}^\top + \boldsymbol{G}, \qquad \boldsymbol{G} \sim \mathsf{GOE}$$

$$\frac{1}{N} D(\boldsymbol{Y} \,\|\, \boldsymbol{G}) \xrightarrow{N\to\infty} \max_{Q \in \mathbb{S}_+^d} \left\{ \mathcal{D}(Q) - \frac{1}{4\lambda}\operatorname{tr}(Q^2) \right\}$$

- Rank-one: Deshpande et al. 2015, Krzakala-Xu 2016, Barbier et al. 2016
- finite rank: Lelarge, Miolane 2017

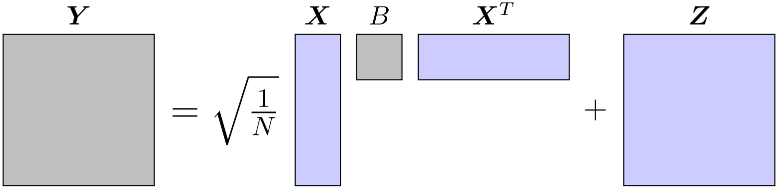
Asymmetric spiked matrix model

$$\boldsymbol{Y} = \sqrt{\frac{\lambda}{N}}\,\boldsymbol{U}\boldsymbol{V}^\top + \boldsymbol{Z}, \qquad \boldsymbol{Z} \overset{\mathsf{IID}}{\sim} \mathsf{N}(0,1)$$

$$\frac{1}{N} D(\boldsymbol{Y} \,\|\, \boldsymbol{Z}) \xrightarrow{N\to\infty} \max_{Q_1 \in \mathbb{S}^d_+} \min_{Q_2 \in \mathbb{S}^d_+} \left\{ \mathcal{D}_1(Q_2) + \mathcal{D}_2(Q_1) - \frac{1}{2\lambda} \operatorname{tr}(Q_1 Q_2) \right\}$$

- Rank-one: Barbier et al. 2017
- finite rank: Miolane 2017

Spiked matrix model with coupling matrix



- Deterministic coupling matrix $B \in \mathbb{R}^{d \times d}$
- Stochastic block model with general interactions **R.**-Mayya-Volvosky 2019

Includes previous models as special cases:

- Symmetric spiked model $\iff$ B is positive (or negative) semidefinite
- Asymmetric spiked model $\iff$ augmentation $+$ asymmetric B

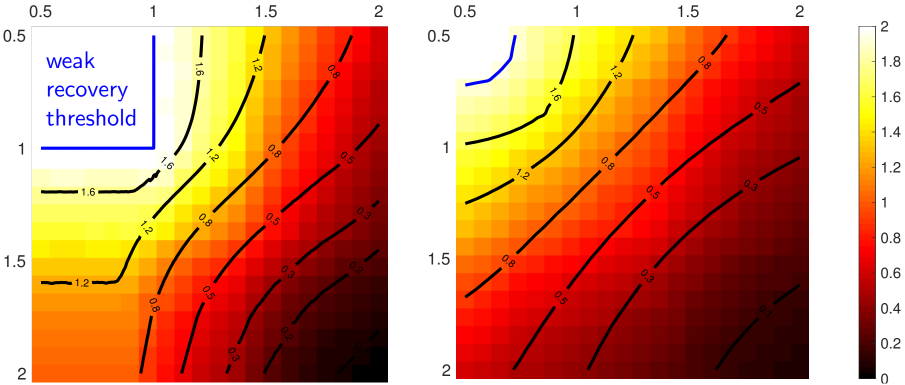
Multiview spiked matrix model

$$\boldsymbol{Y}_\ell = \sqrt{\frac{1}{N}}\,\boldsymbol{X}\,B_\ell\,\boldsymbol{X}^T + \boldsymbol{Z}_\ell \qquad \ell = 1, \dots, L$$

- Applied to community detection with correlated networks by Mayya-**R.** 2019 who proved lower bound on the asymptotic free energy.
- Matching upper bound by Barbier-**R.** 2020 under structural assumption

$$\sum_{\ell=1}^{L} \{(B_\ell \otimes B_\ell) + (B_\ell \otimes B_\ell)^\top\} \succeq 0.$$

Community detection with correlated networks [Mayya-**R.** 2019]



one network $(L = 1)$ two networks $(L = 2)$

Asymptotic MMSE (contour lines) vs. empirical MSE of BP (heat map).
Networks have $N = 10^5$ nodes, 3 communities, and average degree 30.
Axes are eigenvalues of first coupling matrix.

Assumption on the coupling matrices in [Barbier-**R.** 2020] is restrictive

$$\sum_{\ell=1}^{L}\{(B_\ell \otimes B_\ell) + (B_\ell \otimes B_\ell)^\top\} \succeq 0.$$

- Cannot apply to asymmetric model or directed networks
- Formula for asymptotic MMSE is incorrect if assumption is violated.

Is there a universal formula that holds for arbitrary coupling matrices?

Matrix tensor product model

$$\boldsymbol{Y} = \frac{1}{\sqrt{N}}(\boldsymbol{X} \otimes \boldsymbol{X})B + \boldsymbol{Z}$$

- $\boldsymbol{X} \otimes \boldsymbol{X}$ is $N^2 \times d^2$ Kronecker product
- B is $d^2 \times L$ coupling matrix
- Equivalent to multiview model with matrices $B_1, \ldots, B_L$ via

$$B = \begin{bmatrix} \mathrm{vec}(B_1) & \ldots & \mathrm{vec}(B_L) \end{bmatrix}$$

Theorem [**R.** 2020] For any coupling matrix B,

$$\frac{1}{N} D(\boldsymbol{Y} \,\|\, \boldsymbol{Z}) \xrightarrow{N\to\infty} \max_{Q \in \mathbb{S}^d_+} \inf_{R \in \mathbb{S}^d_+} \left\{ \mathcal{D}(R) + \frac{1}{2}\operatorname{tr}(BB^\top (Q \otimes Q)) - \frac{1}{2}\operatorname{tr}(RQ) \right\}$$

Bound on convergence rate is $O(d^4 N^{-1/5})$.

Relation with previous work

- Structural assumption in [Barbier, R. 2020] is equivalent to convexity of

$$Q \mapsto \operatorname{tr}(BB^\top (Q \otimes Q))$$

- If convexity holds, then max-min formula reduces to simplified max formula

$$\begin{aligned}
&\max_{Q \in \mathbb{S}^d_+} \inf_{R \in \mathbb{S}^d_+} \left\{ \mathcal{D}(R) + \frac{1}{2} \operatorname{tr}(BB^\top (Q \otimes Q)) - \frac{1}{2} \operatorname{tr}(RQ) \right\} \\
&= \max_{Q \in \mathbb{S}^d_+} \left\{ \mathcal{D}(T(Q)) - \frac{1}{2} \operatorname{tr}(BB^\top (Q \otimes Q)) \right\}
\end{aligned}$$

where $T \colon \mathbb{S}^d \to \mathbb{S}^d$ is self-adjoint linear operator defined by B.

Why this formula?

Augmented model: Given $(R, S) \in \mathbb{S}^d \times \mathbb{S}^{d^2}$ define

$$\boldsymbol{Y} = \begin{cases} \frac{1}{\sqrt{N}} \boldsymbol{X}^{\otimes 2} S^{1/2} + \boldsymbol{Z}, & \text{MTP model} \\ \boldsymbol{X} R^{1/2} + \boldsymbol{Z}' & \text{linear Gaussian model} \end{cases}$$

$$\mathcal{D}_N(R, S) := \frac{1}{N} D(\boldsymbol{Y} \,\|\, \text{IID Gaussian})$$

Overlap concentration implies matching of derivatives

$$2\nabla_S \mathcal{D}_N = \mathbb{E}\big[(\tfrac{1}{N}\boldsymbol{X}^\top \tilde{\boldsymbol{X}})^{\otimes 2}\big] \approx \big(\mathbb{E}\big[\tfrac{1}{N}\boldsymbol{X}^\top \tilde{\boldsymbol{X}}\big]\big)^{\otimes 2} = \big(2\nabla_R \mathcal{D}_N\big)^{\otimes 2}$$

Heuristic derivation (assuming overlap concentration)

$$\mathcal{D}_N^*(Q,S) := \sup_R \underbrace{\left\{\frac{1}{2}\langle R,Q\rangle - \mathcal{D}_N(R,S)\right\}}_{\text{convex conjugate in 1st arg.}}$$

$$\boxed{\nabla_S \mathcal{D}_N^*(Q,S) = -\nabla_S \mathcal{D}(R^*(Q),S) \overset{\text{overlap conc.}}{\approx} -2(\nabla_R \mathcal{D}(R^*(Q),S))^{\otimes 2} = -\frac{1}{2}Q^{\otimes 2}}$$

$$\mathcal{D}_N^*(Q,S) \approx \mathcal{D}_N^*(Q,0) - \frac{1}{2}\langle S, Q^{\otimes 2}\rangle \qquad (\star)$$

$$\underbrace{\mathcal{D}_N(0,S)}_{\text{original problem}} = \underbrace{\sup_Q\{\langle 0,Q\rangle - \mathcal{D}^*(Q,S)\}}_{\text{biconjugate in 1st arg.}} \overset{(\star)}{\approx} \sup_Q\left\{\langle 0,Q\rangle - \mathcal{D}^*(Q,0) + \frac{1}{2}\langle S, Q^{\otimes 2}\rangle\right\}$$

$$= \sup_Q \inf_R \left\{\mathcal{D}(R) + \frac{1}{2}\langle S, Q^{\otimes 2}\rangle - \langle R,Q\rangle\right\}$$

Main ideas in proof

- Start with adaptive interpolation method [Barbier-Macris 2018]
- Lack of convexity complicates specification of adaptive path
- Rely heavily on order-preserving properties of overlap in Gaussian noise.
- Introduce continuous time variance inequality linking free energy and overlap.

Higher order tensor products?

$$\boldsymbol{Y} = \frac{1}{\sqrt{N^{p-1}}} \boldsymbol{X}^{\otimes p} B + \boldsymbol{Z}$$

Applying result for $p = 2$ recursively suggests following formula

$$\frac{1}{N} D(\boldsymbol{Y} \,\|\, \boldsymbol{Z}) \xrightarrow{N \to \infty} \max_{Q \in \mathbb{S}^d_+} \min_{R \in \mathbb{S}^d_+} \left\{ \mathcal{D}_0(R) + \frac{1}{2} \langle BB^\top, Q^{\otimes p} \rangle - \frac{1}{2} \langle R, Q \rangle \right\}$$

This formula proved via Hamilton–Jacobi equations by Chen-Mourrat-Xia 2021

Table of Contents

Pairwise observations through channel P with input alphabet $\mathcal{X} \times \mathcal{X}$

$$Y_{ij} \overset{\text{ind}}{\sim} P(\cdot \mid x_i, x_j), \quad i, j = 1, \ldots N$$

- What if there are different types of variables?
- What if the observations depend on the variable type?

Groupwise spiked matrix model

- Latent variables partitioned in K groups.
- Obtain pairwise observations for each group

$$\boldsymbol{Y}_{k\ell} = \sqrt{\frac{\lambda_{k\ell}}{N}} \boldsymbol{X}_k \boldsymbol{X}_\ell^\top + \boldsymbol{Z}_{k\ell}, \qquad k, \ell = 1, \dots, K$$

where $\boldsymbol{X}_k$ is $n_k \times 1$ vector of variables in k-th group.

Theorem [Behne-**R.** 2022] If $n_k/N \to \beta_k > 0$ and k-th group IID P_k,

$$\frac{1}{N} D(\boldsymbol{Y} \,\|\, \boldsymbol{Z}) \xrightarrow{N \to \infty} \max_{q \succeq 0} \inf_{r \succeq 0} \left\{ \sum_{k=1}^{K} \mathcal{D}_k(r_k) + \frac{1}{2} \sum_{k,\ell=1}^{K} \lambda_{k\ell}\, q_k q_\ell - \frac{1}{2} r^\top q \right\}$$

- Related recent work by Guionnet-Ko-Krzakala-Zdeborová

Special cases depend on which interactions are observed

$$\Lambda = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1K} \\ \vdots & & \vdots \\ \lambda_{K1} & \dots & \lambda_{KK} \end{pmatrix}$$

- One nonzero (diagonal) $\implies$ symmetric spiked matrix model
- One nonzero (off-diagonal) $\implies$ asymmetric spiked matrix model
- Off-diagonal row $\implies$ generalized spiked covariance model [Bai-Yao 2012]
- Two nonzeros (one diagonal, one off-diagonal) $\implies$ Gaussian version of the contextual stochastic block model (SBM) [Deshpande-Montanari-Mossel-Sen 2018]
- Family of doubly stochastic Toeplitz matrices $\implies$ Proof technique called spatial coupling. Used to study spiked Wigner model [Barbier et al. 2018]
- Tri-diagonal matrix $\implies$ Used to analyze free energy in deep Boltzmann machine with K layers [Albericiet al. 2021]

Implications for weighted PCA

- Heteroskedastic spiked covariance model: Samples from p-variate Gaussian with covariance

$$\Sigma = \sqrt{\frac{\lambda}{N}}\boldsymbol{x}\boldsymbol{x}^\top + \underbrace{\Sigma_0}_{\text{diagonal}}$$

- Overlap of optimally weighted PCA when diagonal entries of Σ_0 are supported on a finite set $\{\sigma_1^2, \ldots, \sigma_L^2\}$, Hong et al. 2018

$$\frac{\langle \boldsymbol{x}, \hat{\boldsymbol{x}} \rangle^2}{\|\boldsymbol{x}\|^2 \|\hat{\boldsymbol{x}}\|^2} \xrightarrow[n \to \infty]{\text{a.s.}} \quad \text{largest real root of } R(z) := 1 - \sum_{\ell=1}^{L} \frac{\beta_\ell}{\sigma_\ell^2} \frac{1-z}{\sigma_\ell^2 + z}$$

Theorem [Behne-**R.** 2022]
The weighted PCA method of Hong et al. 2018 is information-theoretically optimal for a spherical spike.

Table of Contents

Joint with with Riccardo Rossetti

AMP for low-rank + IID Gaussian noise

$$\boldsymbol{Y} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top + \frac{1}{\sqrt{n}}\mathsf{GOE}$$

- Low-rank Lesieur-Krzakala-Zdeborová 2017, Montanari-Venkataramanan 2021
- Non-separable denoisers Berthier-Montanari-Nguyen 2020

Matrix-valued AMP with non-separable denoisers: Gerbelot-Berthier 2021

- Sequence of Lipschitz denoisers: $f_t : \mathbb{R}^{N\times d} \to \mathbb{R}^{N\times d}$
- Initialize $\boldsymbol{X}_0 \in \mathbb{R}^{N\times d}$ and iterate

$$\boldsymbol{M}_t = f_t(\boldsymbol{X}_t), \qquad \boldsymbol{D}_t = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial f_{ti}}{\partial \boldsymbol{x}_{ti}}(\boldsymbol{X}_t)$$

$$\boldsymbol{X}_{t+1} = \boldsymbol{Y}\boldsymbol{M}_t - \boldsymbol{M}_{t-1}\boldsymbol{D}_t$$

State evolution with non-separable denoisers

State Evolution: Define $\boldsymbol{X}_t^* \sim \mathsf{N}(\boldsymbol{X} K_t, \Sigma_t \otimes \mathrm{I}_N)$ recursively via

$$K_0 = \frac{1}{n}\mathbb{E}\big[\boldsymbol{X}^\top f_0(\boldsymbol{X}_0)\big], \qquad \Sigma_0 = \frac{1}{n}\mathbb{E}\big[f_0(\boldsymbol{X}_0)^\top f_0(\boldsymbol{X}_0)\big]$$
$$K_{t+1} = \frac{1}{n}\mathbb{E}\big[\boldsymbol{X}^\top f_t(\boldsymbol{X}_t^*)\big], \qquad \Sigma_{t+1} = \frac{1}{n}\mathbb{E}\big[f_t(\boldsymbol{X}_t^*)^\top f_t(\boldsymbol{X}_t^*)\big]$$

Theorem: Gerbelot-Berthier 2021 Under regularity conditions, for all uniformly pseudo-Lipschitz $\phi_N : \mathbb{R}^{N\times d} \to \mathbb{R}$

$$\Big|\phi_N(\underbrace{\boldsymbol{X}, \boldsymbol{X}_0, \ldots, \boldsymbol{X}_t}_{\text{amp iterations}}) - \mathbb{E}\,\phi_N\,(\underbrace{\boldsymbol{X}, \boldsymbol{X}_0^*, \ldots, \boldsymbol{X}_t^*}_{\text{state evolution}})\Big| \xrightarrow[N\to\infty]{\text{pr}} 0$$

AMP for MTP model

$$\boldsymbol{Y}_\ell = \frac{1}{n}\boldsymbol{X} B_\ell \boldsymbol{X}^\top + \frac{1}{\sqrt{n}}\boldsymbol{Z}_\ell, \quad \ell = 1, \dots, L$$

AMP for MTP: Rossetti-**R.**

- Sequence of Lipschitz denoisers: $f_t : \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times d}$,
- Initialize $\boldsymbol{X}_0 \in \mathbb{R}^{n\times d}$ and iterate

$$\boldsymbol{M}_t = f_t(\boldsymbol{X}_t) \qquad \boldsymbol{D}_t = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial f_{ti}}{\partial \boldsymbol{x}_{ti}}(\boldsymbol{X}_t)$$

$$\boldsymbol{X}_{t+1} = \sum_{\ell=1}^{L}\boldsymbol{Y}_\ell \boldsymbol{M}_t A_{t\ell}^\top + \boldsymbol{Y}_\ell^\top \boldsymbol{M}_t A_{t\ell} - \boldsymbol{M}_{t-1}(A_{t\ell}\boldsymbol{D}_t A_{t\ell}^\top + A_{t\ell}^\top \boldsymbol{D}_t A_{t\ell})$$

Theorem: Optimal reweighting $\iff$ $A_{t\ell} = B_\ell$ for all ℓ, t.

State evolution under optimal reweighting

State Evolution: Define $\boldsymbol{X}_t^* \sim \mathsf{N}(\boldsymbol{X} K_t, \Sigma_t \otimes \mathrm{I}_n)$ recursively via

$$K_0 = \tfrac{1}{n}\boldsymbol{X}^\top f_0(\boldsymbol{X}_0), \qquad \Sigma_0 = \tfrac{1}{n} f_0(\boldsymbol{X}_0^\top) f_0(\boldsymbol{X}_0)$$

$$K_{t+1} = T\big(\tfrac{1}{n}\mathbb{E}\big[\boldsymbol{X}^\top f_t(\boldsymbol{X}_t^*)\big]\big), \qquad \Sigma_{t+1} = T\big(\tfrac{1}{n}\mathbb{E}\big[f_t(\boldsymbol{X}_t^*)^\top f_t(\boldsymbol{X}_t^*)\big]\big)$$

$$T(S) = \sum_\ell B_\ell S B_\ell^\top + B_\ell^\top S B_\ell$$

Linear operator $T \colon \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ is self-adjoint and completely positive

Theorem: SE for Bayes-optimal AMP defined by $\mathbb{S}_+^d$-valued recursion

$$S_{t+1} = T(2\nabla\mathcal{D}(S))$$

Fixed-points $\iff$ Stationary point of IT potential

Table of Contents

- G, Reeves, Information-theoretic limits for the matrix tensor product, *IEEE Journal on Selected Areas in Information Theory*, 2020
- J. K. Behne and G. Reeves, Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity, *AISTATS*, 2023
- R. Rossetti and G. Reeves, Heteroskedastic Low-rank Matrix Factorization [Arxiv, 2023] See poster this evening

Summary and future directions

Pairwise inference model

$$Y_{ij} \overset{\text{ind}}{\sim} P_N(\cdot \mid x_i, x_j), \quad i, j = 1, \dots N$$

MTP model

$$P_N(\cdot \mid x_i, x_j) \quad \approx \quad \mathsf{N}\Big(\sqrt{\tfrac{1}{N}}\, B^\top (x_i \otimes x_j), \mathrm{I}_L\Big), \quad B \in \mathbb{R}^{d^2 \times L}$$

Further directions

- Weak recovery / BBP phase transitions
- Non-Gaussian observations + channel universality
- Spectral methods
- Coupling matrices are unknown, mismatched estimation
- Extrinsic rank, higher-order interactoins