Tensor estimation with structured priors

Clément Luneau, Nicolas Macris

June 29, 2020

Laboratoire de Théorie des Communications, EPFL, Switzerland

Statistical model for tensor estimation

Noisy observations of a symmetric rank-one tensor

$$\forall 1 \leq i \leq j \leq k \leq n : Y_{ijk} = \frac{\sqrt{\lambda}}{n} X_i X_j X_k + Z_{ijk} \qquad \Leftrightarrow \mathbf{Y} = \frac{\sqrt{\lambda}}{n} \mathbf{X}^{\otimes 3} + \mathbf{Z}$$

$$\forall 1 \leq i \leq j \leq n : Y_{ij} = \frac{\sqrt{\lambda}}{n} X_i X_j + Z_{ij} \qquad \Leftrightarrow \mathbf{Y} = \frac{\sqrt{\lambda}}{n} \mathbf{X}\mathbf{X}^T + \mathbf{Z}$$

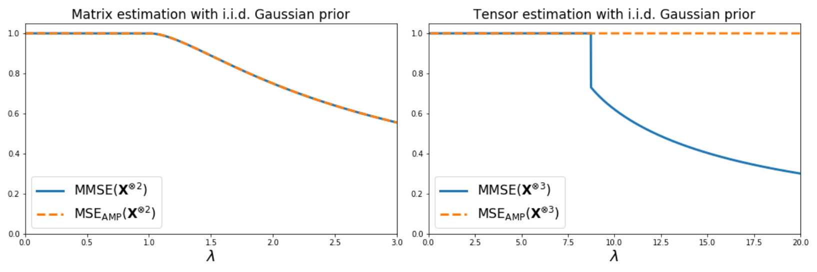
- n-dimensional spike $\mathbf{X} \in \mathbb{R}^n$
- $Z_{ij(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ additive white Gaussian noise
- $\lambda > 0 \propto$ signal-to-noise ratio

Goal: estimate the spike $\mathbf{X}$ and/or the underlying rank-one tensor $\mathbf{X}^{\otimes 3}$

High-dimensional regime for i.i.d. prior

i.i.d. prior on the spike: $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P_X$

- Precise formula[1] for $\mathrm{MMSE} := \frac{1}{n}\mathbb{E}\,\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\|^2$ when $n \to +\infty$
- Performance of Approximate Message Passing algorithm precisely tracked[2]



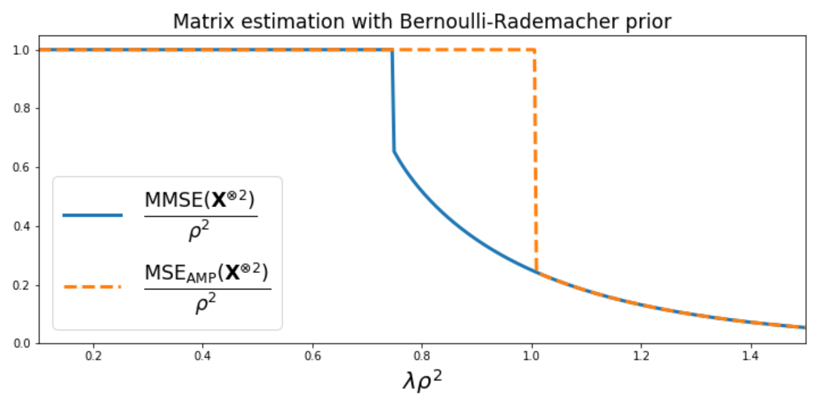
[1] Lelarge and Miolane, "Fundamental limits of symmetric low-rank matrix estimation".
[2] Lesieur et al., "Statistical and computational phase transitions in spiked tensor estimation".

Algorithmic gap for low sparsity prior

Bernoulli-Rademacher prior

$$P_X(1) = P_X(-1) = \rho/2 \quad , \quad P_X(0) = 1 - \rho$$

Algorithmic gap even for matrix estimation if low sparsity ρ (below $\rho = 0.05$)



Structured prior

Data in nature has structure

- Compressed sensing: signal to estimate sparse in some domain
- High-dimensional signal effectively lies on a low-dimensional manifold

Recently[3] use of generative models to encode structure:

$$X_i := \varphi\left(\frac{(\mathbf{W}\mathbf{S})_i}{\sqrt{p}}\right)$$

- $\mathbf{S}$ p-dimensional latent vector: $S_1, \ldots, S_p \overset{\text{i.i.d.}}{\sim} P_S$
- $\mathbf{W}$ sensing matrix: $W_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$
- φ (nonlinear) activation functions

Proposed by Aubin et al. in the context of matrix estimation

[3] Aubin et al., "The spiked matrix model with generative priors".

Matrix estimation with generative priors

High-dimensional limit $n \to +\infty$ with fixed ratio $\alpha := n/p$

"No algorithmic gap with generative-model priors"[4]

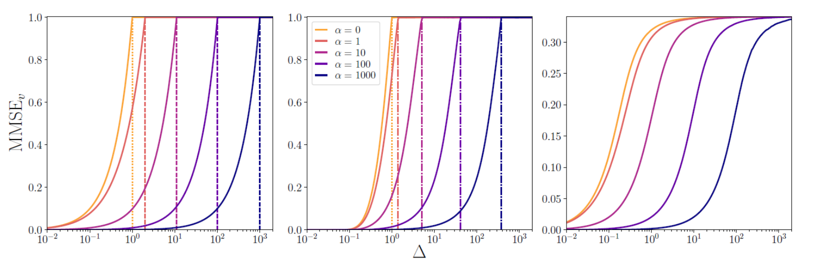


Figure 1: MMSE as a function of $\Delta = 1/\lambda$ for linear (left), sign (centre) and ReLU (right) activations. Figure by Aubin et al.

[4] Aubin et al., "The spiked matrix model with generative priors".

Tensor estimation with generative priors

Can we leverage generative priors in tensor estimation to have a finite algorithmic gap for a centered prior?

In this talk

1. Formulas for asymptotic mutual information & MMSE
2. Visualization of $\mathrm{MMSE}(\mathsf{X}^{\otimes 3})$ for different settings
3. Limit $\alpha := n/p \to 0$: simplified equivalent model with i.i.d. prior

Asymptotic normalized mutual information

$$\forall 1 \leq i \leq j \leq k \leq n : Y_{ijk} = \frac{\sqrt{\lambda}}{n} X_i X_j X_k + Z_{ijk} \text{ with } \forall i : X_i := \varphi\left(\frac{(\mathrm{WS})_i}{\sqrt{p}}\right)$$

Theorem: asymptotic normalized mutual information[5]

$$\lim_{\substack{n \to +\infty \\ n/p \to \alpha}} \frac{I(\mathrm{X};\mathrm{Y}|\mathrm{W})}{n} = \inf_{q_x \in [0,\rho_x]} \inf_{q_s \in [0,\rho_s]} \sup_{r_s \geq 0} \psi_{\lambda,\alpha}(q_x, q_s, r_s)$$

with potential function

$$\begin{aligned}\psi_{\lambda,\alpha}(q_x, q_s, r_s) := & \; I\left(U; \sqrt{\lambda q_x^2/2}\,\varphi(\sqrt{\rho_s - q_s}\,U + \sqrt{q_s}\,V) + \widetilde{Z} \,\middle|\, V\right) \\ & + \frac{1}{\alpha} I(S; \sqrt{r_s}\,S + Z) - \frac{r_s(\rho_s - q_s)}{2\alpha} + \frac{\lambda}{12}(\rho_x - q_x)^2(\rho_x + 2q_x)\end{aligned}$$

where $S \sim P_S$, $U, V, Z, \widetilde{Z} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $\rho_s := \mathbb{E}S^2$, $\rho_x := \mathbb{E}\varphi(\sqrt{\rho_s}U)^2$

[5] Luneau and Macris, Tensor estimation with structured priors.

Minimum mean square error

Theorem: asymptotic tensor MMSE[6]

$$\mathcal{Q}_x^*(\lambda) := \Big\{ q_x^* \in [0,\rho_x] : \inf_{q_s \in [0,\rho_s]} \sup_{r_s \geq 0} \psi_{\lambda,\alpha}(q_x^*, q_s, r_s) = \inf_{q_x \in [0,\rho_x]} \inf_{q_s \in [0,\rho_s]} \sup_{r_s \geq 0} \psi_{\lambda,\alpha}(q_x, q_s, r_s) \Big\}$$

For almost every $\lambda > 0$, $\mathcal{Q}_x^*(\lambda) = \{q_x^*(\lambda)\}$ is a singleton and

$$\lim_{\substack{n \to +\infty \\ n/p \to \alpha}} \frac{\mathbb{E}\left\| \mathbf{X}^{\otimes 3} - \mathbb{E}[\mathbf{X}^{\otimes 3} | \mathbf{Y}, \mathbf{W}] \right\|^2}{n^3} = \rho_x^3 - \big(q_x^*(\lambda)\big)^3$$

[6] Luneau and Macris, Tensor estimation with structured priors.

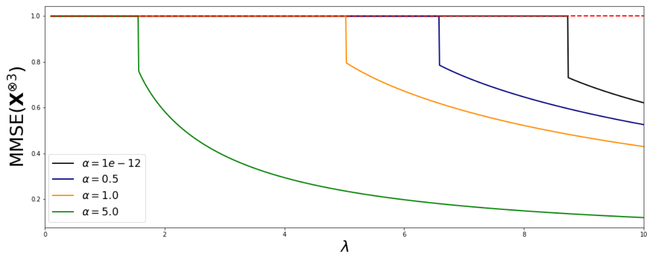
Algorithmic gap

$$\text{critical point equation } \nabla \psi_{\lambda,\alpha}(q_x, q_s, r_s) = 0$$

$$\Updownarrow$$

$$\text{fixed point equation } (q_x, q_s, r_s) = F_{\lambda,\alpha}(q_x, q_s, r_s)$$

- Fixed point with lowest potential $\psi_{\lambda,\alpha}(q_x, q_s, r_s)$ used to compute asymptotic MMSE
- Uninformative fixed point $q_x = 0$ iff φ odd function, P_S centered
  Strongly stable fixed point $\Rightarrow$ infinite algorithmic gap persists



Asymptotic MMSE in the plane (α, λ)

Information theoretic threshold λ_{IT} decreases with the ratio α of signal-to-latent space dimensions

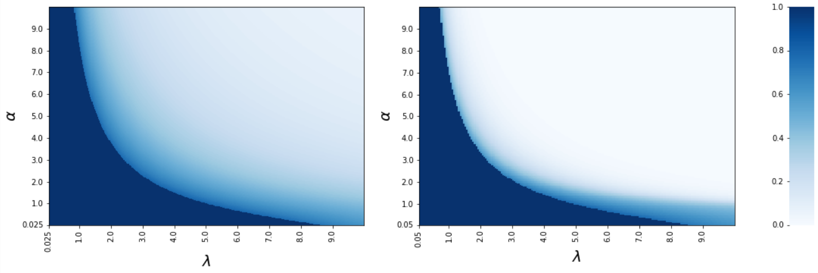


Figure 2: Asymptotic $\mathrm{MMSE}(\mathsf{X}^{\otimes 3})$ as a function of (α, λ) for $\varphi(x) = x$. *Left:* $P_S \sim \mathcal{N}(0, 1)$. *Right:* $P_S \sim \frac{(\delta_1 + \delta_{-1})}{2}$.

Information theoretic threshold λ_{IT} decreases with the ratio α of signal-to-latent space dimensions

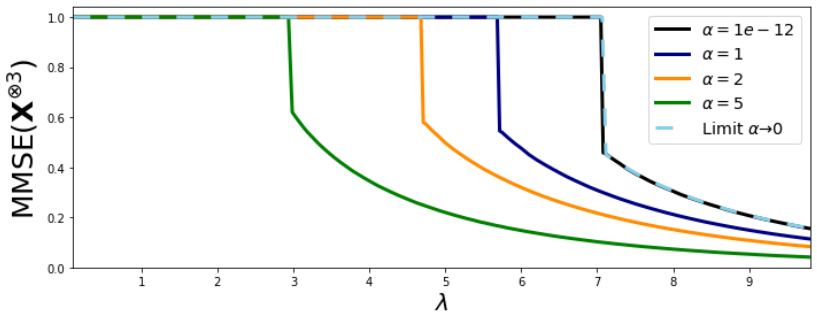


Figure 3: Asymptotic $\mathrm{MMSE}(\mathsf{X}^{\otimes 3})$ as a function of λ for $\varphi(x) = \mathrm{sign}(x)$, $P_S \sim \mathcal{N}(0,1)$ and different values of α. Limit $\alpha \to 0^+$ given by tensor estimation problem with i.i.d. Rademacher prior.

Limit of vanishing signal-to-latent space dimensions

Limit $\alpha \to 0^+$ of the asymptotic mutual information

$$\lim_{\alpha \to 0^+} \lim_{\substack{n \to +\infty \\ n/p \to \alpha}} \frac{I(\mathbf{X};\mathbf{Y}|\mathbf{W})}{n} = \inf_{q_x \in [0,\rho_x]} \frac{\lambda}{12}(\rho_x - q_x)^2(\rho_x + 2q_x) + I\Big(U;\sqrt{\frac{\lambda q_x^2}{2}}\,\varphi\Big(\sqrt{\rho_s - (\mathbb{E}S)^2}\,U + |\mathbb{E}S|V\Big) + \widetilde{Z}\,\Big|V\Big)$$

Same asymptotic mutual information than

$$\widetilde{Y}_{ijk} = \frac{\sqrt{\lambda}}{n}\widetilde{X}_i\widetilde{X}_j\widetilde{X}_k + \widetilde{Z}_{ijk}\ , 1 \leq i \leq j \leq k \leq n\ ,$$

with $\widetilde{X}_i = \varphi(\sqrt{\rho_s - (\mathbb{E}S)^2}\,U_i + |\mathbb{E}S|\,V_i)$; $\mathbf{U},\mathbf{V} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_n)$; <u>$\mathbf{V}$ known</u>

- $\mathbb{E}_{S\sim P_S} S = 0$: i.i.d. prior $\widetilde{X}_1,\dots,\widetilde{X}_n \overset{\text{i.i.d.}}{\sim} \varphi(\mathcal{N}(0,\rho_s))$
- $\mathbb{E}_{S\sim P_S} S \neq 0$: side information $\mathbf{V}$, proof in[7] easily adapted

[7] Lelarge and Miolane, “Fundamental limits of symmetric low-rank matrix estimation”.

Limit of vanishing signal-to-latent space dimensions

"No algorithmic gap with generative-model priors"[8]?

1. Similar behavior for matrix estimation with generative priors
2. We can choose φ to obtain any equivalent i.i.d. prior $\varphi(\mathcal{N}(0, \rho_S))$ when $\alpha \to 0^+$ including a prior exhibiting an algorithmic gap

Algorithmic gap for matrix estimation with generative prior

$\mathsf{X} = \varphi(\mathsf{WS}/\sqrt{p})$ with $S_1, \dots, S_p \overset{\text{i.i.d.}}{\sim} P_S$ centered unit-variance and

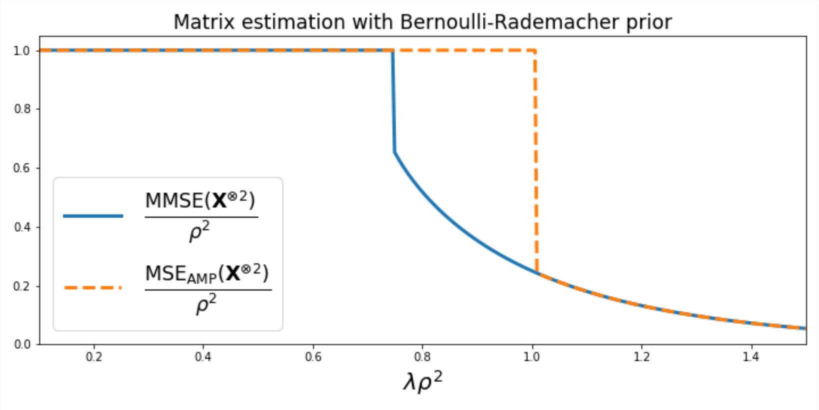
$$\varphi(x) = \begin{cases} -1 \text{ if } x < -\epsilon \\ 0 \quad \text{if } -\epsilon < x < \epsilon \\ +1 \text{ if } x > \epsilon \end{cases} \quad ; \quad \int_{-\infty}^{-\epsilon} \frac{dx}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{\rho}{2}$$

Equivalent to i.i.d. Bernoulli-Rademacher prior when $\alpha \to 0^+$

$$\varphi(\mathcal{N}(0, \rho_S)) \sim (1-\rho)\delta_0 + \frac{\rho}{2}\delta_1 + \frac{\rho}{2}\delta_{-1}$$

[8] Aubin et al., "The spiked matrix model with generative priors".

Limit of vanishing signal-to-latent space dimensions



However regime $\alpha \to 0^+$ does not correspond to a high-dimensional signal $\mathbf{X}$ lying on a lower p-dimensional space

Does the algorithmic gap vanishes/disappears when α increases?

References

- Aubin, Benjamin et al. “The spiked matrix model with generative priors”. In: Advances in Neural Information Processing Systems 32. 2019, pp. 8366–8377.
- Lelarge, Marc and Léo Miolane. “Fundamental limits of symmetric low-rank matrix estimation”. In: Probability Theory and Related Fields 173.3 (2019). ISSN: 1432-2064. DOI: *10.1007/s00440-018-0845-x*.
- Lesieur, Thibault et al. “Statistical and computational phase transitions in spiked tensor estimation”. In: 2017 IEEE International Symposium on Information Theory (ISIT) (2017). DOI: *10.1109/isit.2017.8006580*.
- Luneau, Clément and Nicolas Macris. Tensor estimation with structured priors. 2020. arXiv: *2006.14989 [cs.IT]*.