Information-theoretic limits of a multiview low-rank symmetric spiked matrix model

Jean Barbier* and Galen Reeves†
*The Abdus Salam International Center for Theoretical Physics, Trieste, Italy.
†Department of ECE and Department of Statistical Science, Duke University

***Abstract*—We consider a generalization of an important class of high-dimensional inference problems, namely spiked symmetric matrix models, often used as probabilistic models for principal component analysis. Such paradigmatic models have recently attracted a lot of attention from a number of communities due to their phenomenological richness with statistical-to-computational gaps, while remaining tractable. We rigorously establish the information-theoretic limits through the proof of single-letter formulas for the mutual information and minimum mean-square error. On a technical side we improve the recently introduced adaptive interpolation method, so that it can be used to study low-rank models (i.e., estimation problems of "tall matrices") in full generality, an important step towards the rigorous analysis of more complicated inference and learning models.**

I. Introduction

It is fair to say that in the last decade our rigorous understanding of high-dimensional Bayesian inference has experienced a tremendous improvement. Extremely rich and complex models appearing in a variety of disciplines are now under control, at least at the information-theoretic level, and important steps towards understanding algorithmic limits have been made too. Some paradigmatic examples in high-dimensional regression that have now been put on a rigorous basis are: code division multiple access [1] and sparse superposition codes [2], [3] in communication, that are closely related to compressive sensing in signal processing [1], [4]–[7]; estimation and learning in generalized regression, including as special cases non-linear compressive sensing or the perceptron neural network [8]; or models of shallow and deep neural netwoks [9], [10]. A second class of models that has attracted great attention is matrix (and more generally tensor) estimation, originally introduced as simple probabilistic models for principal component analysis [11]. These papers have generated intense research studying the limits of spectral algorithms [12], semidefinite relaxations and the sum-of-squares hierarchy [13]–[16], the approximate message-passing algorithm (AMP) [17], [18], or gradient-based methods [19].

More closely related to the present paper, the information-theoretic limits of the minimum mean-square error (MMSE) estimator have been studied through a series of results on the validity of so-called "replica symmetric formulas" for the mutual information and MMSE, derived by statistical physics tools [20], and proved by various methods in [21]–[29]. These spiked matrix models are also intimately connected to stochastic block models for community detection [25], [30]–[32].

This work was supported in by the NSF under Grant No. 1750362.

This leap in progress is due in particular to a balanced mixture of new (and older) tools developed mainly in statistical physics and information theory. We mention the rigorous control of AMP through state evolution [33], [34] leading to algorithmic proofs such as [5], [6]; the cavity method [35], [36], applied in inference, e.g., in [25]; interpolation techniques developed in [37], [38] and later improved in the context of inference [27], [39], [40]; information-theoretic proofs [4][1].

The present work is part of this line of research and brings forward two main contributions:

- The introduction and information-theoretic analysis, through the proof of single-letter variational formulas for the mutual information and MMSE, of a generalized spiked matrix model encompassing a number of models in the literature.
- A key improvement of the *adaptive interpolation method* [27], [39], that has proven in the past years to be one of the simplest and most versatile techniques to study high-dimensional inference problems. This novelty allows to generically treat low-rank models that were until now out of reach for the method[2].

II. Multiview symmetric spiked matrix model

A. *Problem formulation*

Let $\boldsymbol{X} = (X_1, \dots, X_n)^\intercal \in \mathbb{R}^{n\times d}$ be an unknown random signal-matrix whose rows are i.i.d. according to a distribution p_X supported on a bounded subset $\mathcal{X}$ of $\mathbb{R}^d$. Consider the observation model:

$$\tilde{\boldsymbol{Y}} = \boldsymbol{X} S^{1/2} + \tilde{\boldsymbol{W}}, \tag{1a}$$

$$\boldsymbol{Y}_\ell = n^{-1/2} \boldsymbol{X} B_\ell \boldsymbol{X}^\intercal + \boldsymbol{W}_\ell, \quad \ell = 1, \dots, L, \tag{1b}$$

where $S \in \mathcal{S}_d^+$, the set of $d \times d$ semidefinite non-negative matrices, and $B_1, \dots, B_L \in \mathbb{R}^{d\times d}$ are deterministically known, while $\tilde{\boldsymbol{W}} \in \mathbb{R}^{n\times d}$ and $\boldsymbol{W}_1, \dots, \boldsymbol{W}_L \in \mathbb{R}^{n\times n}$ are independent noise matrices with i.i.d. standard Gaussian entries. Our analysis considers the regime d, L fixed while taking the limit $n \to +\infty$. We keep the bold notation for matrices with at least one dimension scaling with n.

The observation model (1) encompasses a number of estimation problems found in the literature. The most studied

[1]This reference is also implicitly based on ideas with a flavor of interpolation and cavity methods.

[2]We note the exception of low-rank even-order symmetric tensor estimation [41] that can be treated thanks to non-generic symmetries of the model.

example is the spiked Wigner model when taking $S = 0_d$, $L = 1$ and $B_1 = \gamma \mathrm{I}_d$. Another important one is a class of stochastic blocks models studied in [31], [32].

Note that, by stability of the Gaussian law under addition, this model is equivalent to a model where the first channel in (1) is replaced by K independent Gaussian channels of the form $\tilde{\boldsymbol{Y}}_k = \boldsymbol{X} A_k + \tilde{\boldsymbol{W}}_k$ where $A_k \in \mathbb{R}^{d\times m_k}$. More precisely, this equivalence follows by setting $S = \sum_{k=1}^K A_k A_k^\intercal$, and by equivalence we mean that the Le Cam distance between the two models vanishes [42], which in particular implies mutual information and MMSE-wise equivalence.

One could even further generalize the model by considering additional higher-order tensor observations of the form $Y^{(p)}_{j,i_1\ldots i_p} = n^{\frac{1-p}{2}} \sum_{k_1,\ldots,k_p}^d C^{(p)}_{j,k_1\ldots k_p} X_{i_1k_1}\ldots X_{i_pk_p} + W^{(p)}_{j,i_1\ldots i_p}$ with $C_j^{(p)}$ an order-p coupling tensor and $\boldsymbol{W}_j^{(p)}$ an order-p Gaussian noise tensor, with $p \geq 3$, $1 \leq j \leq J$, $(i_1,\ldots,i_p) \in \{1,\ldots,n\}^p$. The analysis would be more involved notation-wise but would not present any additional difficulties, and is left for future work.

B. The minimum mean-square error matrix

Our results describe the information-theoretic limits of estimating the signal matrix $\boldsymbol{X}$ from the collection of observations $\boldsymbol{Y} \equiv (\boldsymbol{Y}_\ell)$ and $\tilde{\boldsymbol{Y}}$. We focus on the asymptotic mutual information as well as a multivariate performance measure called the MMSE matrix [31], which is defined by

$$\mathsf{MMSE}(\boldsymbol{X} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}) \equiv \tfrac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathsf{Cov}(X_i \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y})],$$

where the $d\times d$ conditional covariance matrix $\mathsf{Cov}(X_i \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y})$ of the row X_i of $\boldsymbol{X}$ (which, when considered alone, is a column vector) is, more explicitly,

$$\mathbb{E}\big[(X_i - \mathbb{E}[X_i \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}])(X_i - \mathbb{E}[X_i \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}])^\intercal \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}\big].$$

Note that the trace of this matrix corresponds to the usual definition of the MMSE:

$$\mathrm{Tr}\,\mathsf{MMSE}(\boldsymbol{X} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}) = \tfrac{1}{n}\mathbb{E}\big[\|\boldsymbol{X} - \mathbb{E}[\boldsymbol{X} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}]\|_{\mathrm{F}}^2\big].$$

One of the advantages of the matrix MMSE is that is provides information about the individual dimensions in the row space of $\boldsymbol{X}$, which can be important in settings where some dimensions can be recovered while other cannot; see [10], [31].

Our analysis leverages the matrix I-MMSE relation [43], [44], which relates the MMSE matrix to the gradient of the mutual information:

$$\mathsf{MMSE}(\boldsymbol{X} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}) = \tfrac{2}{n}\nabla_S I\big(\boldsymbol{X}; (\tilde{\boldsymbol{Y}}, \boldsymbol{Y})\big),$$

where S is the $d\times d$ matrix appearing in (1a). Note that evaluating this gradient at $S = 0_d$ recovers the MMSE matrix associated with only the observations $\boldsymbol{Y}$.

C. Statement of main results

We provide formulas that depend only on the mutual information and MMSE associated with a d-dimensional estimation problem. Define for $S \in \mathcal{S}_d^+$, the functions

$$I_S : r \in \mathcal{S}_d^+ \mapsto I(X; (S+r)^{1/2}X + W)$$
$$M_S : r \in \mathcal{S}_d^+ \mapsto \mathbb{E}[\mathsf{Cov}(X \mid (S+r)^{1/2}X + W)],$$

where $X \sim p_X$, $W \sim \mathcal{N}(0, \mathrm{I}_d)$. Let $\bar{\rho} \equiv \mathbb{E}[XX^\intercal]$ be the $d\times d$ second moment matrix of p_X and define $f : \mathcal{S}_d^+ \mapsto \mathbb{R}$ as

$$f(S) \equiv \inf_{r\in\mathcal{S}_d^+} \sup_{q\in\mathcal{S}_d^+} \mathcal{I}(r,q) \tag{2}$$

where the following "replica symmetric potential" function $\mathcal{I}(r,q)$ depends also implicitly on $(S, (B_\ell))$:

$$\mathcal{I}(r,q) \equiv I_S(r) + \tfrac{1}{2}\mathrm{Tr}\big[r(q-\bar{\rho}) + \textstyle\sum_\ell^L \{B_\ell^\intercal \bar{\rho} B_\ell \bar{\rho} - B_\ell^\intercal q B_\ell q\}\big].$$

Our results require some structure for the (B_ℓ) matrices:

Hypothesis 1 (Positive coupling structure)**.** *The (B_ℓ) are s.t.*

$$\textstyle\sum_{\ell=1}^L \{(B_\ell \otimes B_\ell) + (B_\ell \otimes B_\ell)^\intercal\} \succcurlyeq 0.$$

Theorem II.1 (Replica symmetric formulas)**.** *Under hypothesis 1 the mutual information for model* (1) *verifies,*

$$\lim_{n\to+\infty} \tfrac{1}{n} I\big(\boldsymbol{X}; (\tilde{\boldsymbol{Y}}, \boldsymbol{Y})\big) = f(S) \tag{3}$$

for all $S \in \mathcal{S}_d^+$. Furthermore, f is concave and hence differentiable almost everywhere on $\mathcal{S}_d^{++}$. At each point $S \in \mathcal{S}_d^{++}$ where f is differentiable, the MMSE matrix verifies

$$\lim_{n\to+\infty} \mathsf{MMSE}(\boldsymbol{X} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}) = \tfrac{1}{2}\nabla f(S). \tag{4}$$

Note that this theorem implies in particular the tightness of the bounds for community detection provided in [31], [32].

Let us comment briefly on the motivation for including the submodel (1a) in our analysis. Many of the problem formulations in the literature correspond to the setting $S = 0_d$. In these settings, our results characterize the exact limit of the mutual information, and using standard perturbation arguments similar to [8], [25], one can then verify the limiting behavior of the scalar MMSE associated with each $n \times n$ matrix $\boldsymbol{X} B_\ell \boldsymbol{X}^\intercal$.

However, if the goal it to recover the matrix $\boldsymbol{X}$ itself, then the analysis becomes more difficult due to invariances in the problem. For example, if $\boldsymbol{X}$ is equal in distribution to $-\boldsymbol{X}$ and the matrices (B_ℓ) are symmetric, then the posterior distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is also symmetric with respect to a sign change. In this case the conditional expectation is deterministically zero and so the MMSE is constant. As is argued in [31] the inclusion of submodel (1a) provides an approach to resolve these non-identifiability issues that is both interpretive and intuitive. The basic idea is that an arbitrarily small but positive definite S is sufficient to break the symmetry in the model and thus the double limit

$$\lim_{S\to 0_d} \lim_{n\to+\infty} \mathsf{MMSE}(\boldsymbol{X} \mid \tilde{\boldsymbol{Y}}, \boldsymbol{Y}),$$

provides a meaningful measure of performance, even if the $S = 0_d$ limit is degenerate.

The potential function $\mathcal{I}$ verifies the following stationary conditions, called *state evolution* equations:

$$\begin{aligned} \nabla_r \mathcal{I}(r,q) = 0_d &\Leftrightarrow q = \bar{\rho} - M_S(r), \\ \nabla_q \mathcal{I}(r,q) = 0_d &\Leftrightarrow r = \textstyle\sum_{\ell=1}^L \big\{B_\ell q B_\ell^\intercal + B_\ell^\intercal q B_\ell\big\} \\ &\qquad \equiv r^*(q). \end{aligned} \tag{5}$$

At a stationary point with respect to q the potential is

$$\mathcal{I}(r^*(q),q)=I_S(r^*(q))+\tfrac{1}{2}\sum_\ell^L \mathrm{Tr}\big[B_\ell^\intercal(\bar\rho-q)B_\ell(\bar\rho-q)\big]. \quad (6)$$

A fact we will need later in the proof, and which is the reason for hypothesis 1, is the following:

Lemma II.1 (Concavity of potential)**.** *Under hypothesis 1 the potential $q\in\mathcal{S}_d^+\mapsto\mathcal{I}(r,q)$ is concave.*

Proof. Using vectorization, the q-dependent piece of the potential $g : q\in\mathcal{S}_d^+ \mapsto \mathrm{Tr}[rq]-\sum_\ell \mathrm{Tr}[B_\ell^\intercal q B_\ell q]$ can be expressed as

$$\begin{aligned} g(q) = \mathsf{vech}(q) D_d^\intercal D_d\,\mathsf{vech}(r)\\ - \mathsf{vech}(q)^\intercal \textstyle\sum_\ell D_d^\intercal (B_\ell\otimes B_\ell) D_d\,\mathsf{vech}(q)\end{aligned}$$

where $\mathsf{vech}(q)$ is the $d(d+1)/2\times 1$ vector obtained by stacking the entries on or below the diagonal of q, and D_d is the duplication matrix, i.e., the unique $d^2\times d(d+1)/2$ matrix such that for any $M\in\mathcal{S}_d$ (the set of symmetric $d\times d$ matrices), $\mathsf{vec}(M)=D_d\,\mathsf{vech}(M)$ where $\mathsf{vec}(M)$ is the $d^2\times 1$ vector obtained by stacking the columns of M; see [45, Chapter 3.8]. Thus, the Hessian of g can be expressed as $-D_n^\intercal\sum_\ell\{(B_\ell\otimes B_\ell)+(B_\ell\otimes B_\ell)^\intercal\}D_n$. Under hypothesis 1, this matrix is negative semidefinite and therefore the concavity of the potential follows. □

Formula (2) can also be expressed in terms of a one-letter potential. This expression is equivalent[3] to the one given in [32, Theorem 1] for the case where (B_ℓ) are symmetric.

Lemma II.2 (Single-Letter Formula)**.** *The function $f(S)$ in* (2) *can also be expressed as*

$$\begin{aligned} f(S)=\inf_{q\in\mathcal{S}_d^+}\Big\{ & I_S\big(\textstyle\sum_{\ell=1}^L\{B_\ell q B_\ell^\intercal + B_\ell^\intercal q B_\ell\}\big)\\ & + \tfrac12\mathrm{Tr}\big[\textstyle\sum_{\ell=1}^L\{B_\ell^\intercal(\bar\rho-q)B_\ell(\bar\rho-q)\}\big]\Big\}.\end{aligned}$$

Proof. Let $r^*(q)$ be given by (5). For each $\tilde q$ it follows Lemma II.1 that $\sup_{q\in\mathcal{S}_d^+}\mathcal{I}(r^*(\tilde q),q)=\mathcal{I}(r^*(\tilde q),\tilde q)$. Restricting the infimum over r to the image of $q\mapsto r^*(q)$ leads to an upper bound

$$f(S)\le \inf_{\tilde q\in\mathcal{S}_d^+}\sup_{q\in\mathcal{S}_d^+}\mathcal{I}(r^*(\tilde q),q)=\inf_{\tilde q\in\mathcal{S}_d^+}\mathcal{I}(r^*(\tilde q),\tilde q).$$

Alternatively, because $\mathcal{I}(r,q)$ is convex-concave, standard duality arguments show that we can interchange the role of r and q in the inf sup, and this leads to a matching lower bound

$$f(S)=\inf_{q\in\mathcal{S}_d^+}\sup_{r\in\mathcal{S}_d^+}\mathcal{I}(r,q)\ge\inf_{q\in\mathcal{S}_d^+}\mathcal{I}(r^*(q),q).$$

Comparing with (6) this gives the stated result. □

III. Adaptive Interpolation, Reloaded

In this section we prove the first part of theorem II.1 using an evolution of the adaptive interpolation method [27], [39].

A. Interpolating model

We consider a model parameterized by the *time* $t\in[0,1]$, a *perturbation*

$$\epsilon=(\lambda,\eta\,\mathrm{I}_d)\in\mathcal{S}_d^+\times\mathcal{S}_d^+$$

with a scalar $\eta\ge 0$, and a generic *interpolation function* $R(t,\eta)\in\mathcal{S}_d^+$ verifying $R(0,\eta)=0$. Note that the interpolation function depends on η but *not* on λ. We require that $\epsilon=\epsilon_n$ verifies $\|\epsilon\|_\mathrm{F}\le b_n\to 0_+$. Define $\boldsymbol{X}_\ell\equiv\boldsymbol{X}B_\ell\boldsymbol{X}^\intercal$. The *interpolating model* is

$$\tilde{\boldsymbol{Y}}(t,\eta)=\boldsymbol{X}\big(S+\eta\,\mathrm{I}_d+R(t,\eta)\big)^{1/2}+\tilde{\boldsymbol{W}}, \quad (7a)$$

$$\hat{\boldsymbol{Y}}(\lambda)=\boldsymbol{X}\lambda^{1/2}+\hat{\boldsymbol{W}}, \quad (7b)$$

$$\boldsymbol{Y}_\ell(t)=\sqrt{(1-t)/n}\,\boldsymbol{X}_\ell+\boldsymbol{W}_\ell,\quad \ell=1,\dots,L. \quad (7c)$$

where all the $\boldsymbol{W}$ matrices are independent and made of i.i.d. $\mathcal{N}(0,1)$ entries. Let $\boldsymbol{Y}(t)\equiv(\boldsymbol{Y}_\ell(t))$. The *interpolating mutual information* is

$$\mathcal{I}_n(t,\epsilon)\equiv\tfrac1n I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t))\big).$$

By construction it verifies

$$\begin{aligned}\mathcal{I}_n(0,\epsilon)&=\tfrac1n I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big)+O(b_n),\\ \mathcal{I}_n(1,\epsilon)&=I_S(R(1,\eta))+O(b_n).\end{aligned}$$

So at $t=0$ the mutual information of the original model (1) appears naturally (recall $\boldsymbol{Y}\equiv(\boldsymbol{Y}_\ell)$). The $O(b_n)$'s, that are uniform in (t,ϵ), are extracted using the chain rule for mutual information and the Lipschitzianity of $\epsilon\mapsto\mathcal{I}_n(t,\epsilon)$, with Lipschitz constant depending only on $|\sup\mathcal{X}|<\infty$ (this follows from the I-MMSE relation, see [29] for similar computations). We compare these values using the fundamental theorem of calculus

$$\mathcal{I}_n(0,\epsilon)=\mathcal{I}_n(1,\epsilon)-\textstyle\int_0^1\mathcal{I}_n'(t,\epsilon)dt,$$

where the prime $'$ will always mean t-derivative. Denote the expectation with respect to the interpolating model posterior $P(\cdot\,|\,\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t))$ as

$$\langle-\rangle_{t,\epsilon}\equiv\mathbb{E}[-|\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t)].$$

We evaluate $\mathcal{I}_n'$ which is directly obtained from the matrix I-MMSE relation. Denote $\tilde{\boldsymbol{X}}\in\mathcal{X}^{n\times d}$ a random sample from the posterior $P(\cdot\,|\,\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t))$, the concise notation $\tilde{\boldsymbol{X}}_\ell\equiv\tilde{\boldsymbol{X}}B_\ell\tilde{\boldsymbol{X}}^\intercal$, and finally $\mathbb{E}$ the joint expectation with respect to the data $(\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t))$. Define

$$\rho\equiv\tfrac1n\boldsymbol{X}^\intercal\boldsymbol{X}\qquad\text{and}\qquad Q\equiv\tfrac1n\boldsymbol{X}^\intercal\tilde{\boldsymbol{X}},$$

(recall $\bar\rho\equiv\mathbb{E}[XX^\intercal]=\mathbb{E}\rho$) where Q is the $d\times d$ *overlap* matrix. Then the mutual information time derivative reads

$$\begin{aligned}2\,\mathcal{I}_n'(t,\epsilon)&=\mathbb{E}\,\mathrm{Tr}\big[\tfrac1n R'(t,\eta)(\boldsymbol{X}-\langle\tilde{\boldsymbol{X}}\rangle_{t,\epsilon})^\intercal(\boldsymbol{X}-\langle\tilde{\boldsymbol{X}}\rangle_{t,\epsilon})\big]\\ &\quad-\textstyle\sum_\ell\mathbb{E}\,\mathrm{Tr}\big[\tfrac{1}{n^2}(\boldsymbol{X}_\ell-\langle\tilde{\boldsymbol{X}}_\ell\rangle_{t,\epsilon})^\intercal(\boldsymbol{X}_\ell-\langle\tilde{\boldsymbol{X}}_\ell\rangle_{t,\epsilon})\big]\\ &=\mathrm{Tr}\big[R'(t,\eta)(\bar\rho-\mathbb{E}\langle Q\rangle_{t,\epsilon})\big]\\ &\quad-\textstyle\sum_\ell\mathrm{Tr}\big[\mathbb{E}[B_\ell\rho B_\ell^\intercal\rho]-\mathbb{E}\langle B_\ell Q^\intercal B_\ell^\intercal Q\rangle_{t,\epsilon}\big],\end{aligned}$$

[3]There is factor of two difference in the formulas that arises from the fact that [32] considers symmetric noise matrices.

using that the average overlap $\mathbb{E}\langle Q\rangle_{t,\epsilon}$ is a symmetric matrix, as well as the two identities [4]

$$\mathbb{E}[\boldsymbol{X}^\intercal\langle\tilde{\boldsymbol{X}}\rangle_{t,\epsilon}] = \mathbb{E}[\langle\tilde{\boldsymbol{X}}\rangle_{t,\epsilon}^\intercal\langle\tilde{\boldsymbol{X}}\rangle_{t,\epsilon}]$$
$$\text{and}\quad \mathbb{E}[\boldsymbol{X}_\ell^\intercal\langle\tilde{\boldsymbol{X}}_\ell\rangle_{t,\epsilon}] = \mathbb{E}[\langle\tilde{\boldsymbol{X}}_\ell\rangle_{t,\epsilon}^\intercal\langle\tilde{\boldsymbol{X}}_\ell\rangle_{t,\epsilon}].$$

Plugging this back in the fundamental theorem of calculus above, using that $\rho = \bar{\rho} + o_n(1)$ by the law of large numbers, and having in mind that we will soon exploit a concentration property for Q, we obtain the *sum rule*:

$$\begin{aligned}\tfrac{1}{n}I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big) &= I_S(R(1,\eta)) + o_n(1) \qquad (8)\\ &+ \tfrac{1}{2}\textstyle\int_0^1 dt\big\{\mathrm{Tr}\big[R'(t,\eta)(\mathbb{E}\langle Q\rangle_{t,\epsilon}-\bar{\rho})\big] + \mathcal{R}^{\mathrm{fluc}}(t,\epsilon)\\ &+ \textstyle\sum_\ell \mathrm{Tr}\big[B_\ell^\intercal\bar{\rho}B_\ell\bar{\rho} - B_\ell^\intercal\mathbb{E}\langle Q\rangle_{t,\epsilon}B_\ell\mathbb{E}\langle Q\rangle_{t,\epsilon}\big]\big\},\end{aligned}$$

where $o_n(1)$ vanishes uniformly in (t,ϵ) and the *remainder* is

$$\mathcal{R}^{\mathrm{fluc}}(t,\epsilon) \equiv \textstyle\sum_{\ell=1}^L \mathrm{Tr}\big[B_\ell\mathbb{E}\big\langle Q^\intercal B_\ell^\intercal(\mathbb{E}\langle Q\rangle_{t,\epsilon}-Q)\big\rangle_{t,\epsilon}\big].$$

This remainder is small if the overlap fluctuations w.r.t. the measure $\mathbb{E}\langle-\rangle_{t,\epsilon}$ are small. This is shown as follows. Consider a perturbation $\lambda\in\mathcal{D}_n \equiv \{\lambda\in\mathcal{S}_d^+ : \lambda_{kk'}\in(s_n,2s_n)$ if $k'\neq k, \lambda_{kk}\in(2ds_n,(2d+1)s_n)\}$ for a sequence (s_n) of positive numbers that accumulate to 0. Because the interpolating model depends on λ *only* through the explicit dependence in channel (7b), the results of [46] *directly apply*. There it is shown that quite generically the overlap concentrates under $\mathbb{E}\langle-\rangle_{t,\epsilon}$, in λ-average, as long as the "free energy" F_n concentrates:

$$F_n \equiv -\tfrac{1}{n}\ln\textstyle\int dP_X(\tilde{\boldsymbol{x}})P(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t))\,.$$

Notice that $\mathbb{E}F_n = \frac{1}{n}h(\tilde{\boldsymbol{Y}}(t,\eta),\hat{\boldsymbol{Y}}(\lambda),\boldsymbol{Y}(t))$ is the differential entropy of the data.

Proposition III.1 (Overlap concentration [46])**.** *Suppose*

$$\mathbb{E}\big[(F_n-\mathbb{E}F_n)^2\big] = O(\tfrac{1}{n}).$$

Let $\mathbb{E}_\lambda[-]\equiv\mathrm{Vol}(\mathcal{D}_n)^{-1}\int_{\mathcal{D}_n}d\lambda[-]$. *Then we have, uniformly in* (t,η) *and the choice of* R,

$$\mathbb{E}_\lambda\mathbb{E}\big\langle\|Q-\mathbb{E}\langle Q\rangle_{t,\epsilon}\|_{\mathrm{F}}^2\big\rangle_{t,\epsilon} = O((s_n^4 n)^{-1/6}).$$

Proving $\mathbb{E}[(F_n-\mathbb{E}F_n)^2] = O(\frac{1}{n})$ whenever $P_X = p_X^{\otimes n}$, i.e., decouples over the rows, is standard. E.g., one can slightly adapt the proof of concentration found in [29], [39]. Equipped with proposition III.1 a simple application of the Cauchy-Schwarz inequality yields $\mathbb{E}_\lambda|\mathcal{R}^{\mathrm{fluc}}(t,\epsilon)| = o_n(1)\to 0_+$ uniformly in (t,η) as $n\to+\infty$, when choosing an appropriate sequence (s_n) that vanishes (but not too fast, i.e., $s_n=\omega(n^{-1/4})$). Then by Fubini's theorem we have, uniformly in η and R,

$$\mathbb{E}_\lambda\textstyle\int_0^1 dt\,\mathcal{R}^{\mathrm{fluc}}(t,\epsilon) = o_n(1). \qquad (9)$$

[4] Such identities are sometimes called "Nishimori identities" in the literature.

B. Upper bound on mutual information

We now exploit our freedom of choice of the interpolation function R and set its derivative to a constant $R'(t,\eta) = r\in\mathcal{S}_d^+$, so that $R(1,\eta)=r$ too. Under this choice, averaging the sum rule (8) w.r.t. $\lambda\in\mathcal{D}_n$ and using (9) yields

$$\tfrac{1}{n}I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big) = \mathbb{E}_\lambda\textstyle\int_0^1 dt\,\mathcal{I}(r,\mathbb{E}\langle Q\rangle_{t,\epsilon}) + o_n(1)\ \forall\ r\in\mathcal{S}_d^+.$$

Therefore we obtain the upper bound:

$$\limsup_{n\to+\infty}\frac{1}{n}I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big) \le \inf_{r\in\mathcal{S}_d^+}\sup_{q\in\mathcal{S}_d^+}\mathcal{I}(r,q). \qquad (10)$$

C. Lower bound on mutual information

Let a new perturbation

$$\tilde{\epsilon}\equiv(0_d,\eta\,\mathrm{I}_d).$$

This time we select R as the unique solution of the following differential equation with initial condition $R(0,\eta)=0$:

$$\begin{aligned}R'(t,\eta) &= \textstyle\sum_{\ell=1}^L\big\{B_\ell\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}B_\ell^\intercal + B_\ell^\intercal\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}B_\ell\big\}\\ &= r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}). \qquad (11)\end{aligned}$$

We used definition (5) for r^*. With the notation $\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}$ we emphasize that we consider an expectation of the overlap *along an interpolation path in which we set the perturbation* λ *to the all zeros matrix* 0_d *while the other perturbation* $\eta\,\mathrm{I}_d$ *is unchanged*. The expected overlap $\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}$ is a function of $R(t,\eta)$ so the equation above is an ODE. $\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}$ is (component-wise) Lipschitz in R with Lipschitz constant depending only on $|\sup\mathcal{X}|$ and n that are both finite (this is seen using relations found, e.g., in [44]), so by the Cauchy-Lipschitz theorem this first order ODE admits a unique global $\mathcal{C}^1$ solution $R_n^*(t,\eta)$, which importantly is independent of λ. Using this solution the derivative of the mutual information is

$$\begin{aligned}2\,\mathcal{I}_n'(t,\epsilon) = &-\textstyle\sum_\ell^L\mathrm{Tr}\big[B_\ell^\intercal(\bar{\rho}-\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})B_\ell(\bar{\rho}-\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})\big]\\ &+\mathcal{R}^{\mathrm{mis}}(t,\epsilon)-\mathcal{R}^{\mathrm{fluc}}(t,\epsilon),\end{aligned}$$

where a new remainder appeared (here $\times$ is the matrix product):

$$\begin{aligned}\mathcal{R}^{\mathrm{mis}}(t,\epsilon)\equiv\textstyle\sum_\ell^L\mathrm{Tr}\big[&B_\ell^\intercal(\mathbb{E}\langle Q\rangle_{t,\epsilon}-\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})B_\ell\\ &\times(\mathbb{E}\langle Q\rangle_{t,\epsilon}-\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})\big].\end{aligned}$$

The remainder $|\mathcal{R}^{\mathrm{mis}}(t,\epsilon)|$ is small if the mismatch $|\mathbb{E}\langle Q\rangle_{t,\epsilon}-\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}|$ due to the different choices of interpolation paths (that differ in the perturbation λ) is small. The purpose of the other perturbation $\eta\,\mathrm{I}_d$, which is a novelty w.r.t. the usual adaptive interpolation method, is to control this remainder:

Proposition III.2 (Interpolation paths mismatch)**.** *Let* $\delta>0$, $\lambda\in\mathcal{D}_n$ *and* $\mathbb{E}_\eta[-]\equiv\frac{1}{\delta}\int_0^\delta d\eta[-]$. *Then, uniformly in* (λ,R),

$$\mathbb{E}_\eta\textstyle\int_0^1 dt\,\mathcal{R}^{\mathrm{mis}}(t,\epsilon) = O(s_n/\delta).$$

Proof. Notice $\mathbb{E}\langle Q\rangle_{t,\epsilon}-\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}\succcurlyeq 0_d$ or, said differently, the matrix MMSE is a decreasing function (w.r.t. the Loewner partial order) of the signal-to-noise matrix λ [47, Prop. 5]. Using Cauchy-Schwarz and $\mathrm{Tr}A \geq \|A\|_{\mathrm{F}}$ for $A \succcurlyeq 0_d$,

$$|\mathcal{R}^{\mathrm{mis}}(t,\epsilon)| \leq \|B_\ell^\intercal(\mathbb{E}\langle Q\rangle_{t,\epsilon} - \mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})B_\ell\|_{\mathrm{F}} \times \mathrm{Tr}[\mathbb{E}\langle Q\rangle_{t,\epsilon} - \mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}]. \tag{12}$$

Now notice that by the I-MMSE matrix relation [44] we have $\mathrm{Tr}[(\mathrm{I}_d + \frac{d}{d\eta}R_n^*(t,\eta))(\mathbb{E}\langle Q\rangle_{t,\epsilon} - \mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})] = 2\frac{d}{d\eta}(\mathcal{I}_n(t,\tilde{\epsilon}) - \mathcal{I}_n(t,\epsilon))$. Moreover we have $\frac{d}{d\eta}R_n^*(t,\eta) \succcurlyeq 0_d$. This comparison inequality is obtained using a comparison inequality similar to [48, Lemma 5]. We refer to [49] where this step is proven in full details

As the trace of a product of non-negative definite matrices is non-negative,

$$\mathrm{Tr}[\mathbb{E}\langle Q\rangle_{t,\epsilon} - \mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}] \leq 2\tfrac{d}{d\eta}(\mathcal{I}_n(t,\tilde{\epsilon}) - \mathcal{I}_n(t,\epsilon)).$$

Therefore averaging (12) over $\eta \in (0,\delta)$ gives

$$\begin{aligned}\mathbb{E}_\eta|\mathcal{R}^{\mathrm{mis}}(t,\epsilon)| &\leq \tfrac{C}{\delta}\int_0^\delta d\eta \tfrac{d}{d\eta}(\mathcal{I}_n(t,\tilde{\epsilon}) - \mathcal{I}_n(t,\epsilon)) \\ &= \tfrac{C}{\delta}\big\{\mathcal{I}_n\big(t,(0_d,\delta\,\mathrm{I}_d)\big) - \mathcal{I}_n\big(t,(\lambda,\delta\,\mathrm{I}_d)\big) \\ &\quad - \mathcal{I}_n\big(t,(0_d,0_d)\big) + \mathcal{I}_n\big(t,(\lambda,0_d)\big)\big\}\end{aligned}$$

for some constant $C > 0$ dependent only on $|\sup\mathcal{X}|$ and (B_ℓ). By Lipschitzianity of $\lambda \mapsto \mathcal{I}_n(t,(\lambda,A))$, with Lipschitz constant depending only on $|\sup\mathcal{X}|$, we get $\mathbb{E}_\eta|\mathcal{R}^{\mathrm{mis}}(t,\epsilon)| = O(\|\lambda\|_{\mathrm{F}}/\delta)$. Finally using $\|\lambda\|_{\mathrm{F}} = O(s_n)$ and Fubini's theorem to switch t and η averages ends the proof. □

The function $I_S(R)$ is concave in $\mathcal{S}_d^+$ [43], [44]. Then, because $R(1,\eta) = \int_0^1 dt\, R'(t,\eta)$,

$$I_S(R(1,\eta)) \geq \textstyle\int_0^1 dt\, I_S(R'(t,\eta)).$$

Let $\delta_n = o_n(1)$ such that $s_n/\delta_n = o_n(1)$. With the above inequality combined with proposition III.2 and (9) (and the uniformity of the bounds), averaging the sum rule (8) with respect to $(\lambda,\eta) \in \mathcal{D}_n \times (0,\delta_n)$ gives (recall definition (5))

$$\begin{aligned}\tfrac{1}{n}I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big) &\geq \mathbb{E}_{\eta,\lambda}\textstyle\int_0^1 dt\big\{I_S(r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})) \\ &\quad + \tfrac{1}{2}\textstyle\sum_\ell^L \mathrm{Tr}\big[B_\ell^\intercal(\bar{\rho} - \mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})B_\ell(\bar{\rho} - \mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}})\big]\big\} + o_n(1) \\ &= \mathbb{E}_{\eta,\lambda}\textstyle\int_0^1 dt\,\mathcal{I}\big(r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}),\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}\big) + o_n(1)\end{aligned}$$

where we used (6) in order to identify the potential. Thanks to our choice of interpolation path (11) we have the identity

$$\mathcal{I}\big(r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}),\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}\big) = \sup\nolimits_{q\in\mathcal{S}_d^+}\mathcal{I}\big(r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}),q\big).$$

Indeed, by lemma II.1 the function $q \in \mathcal{S}_d^+ \mapsto \mathcal{I}(r,q)$ is concave under hypothesis 1. The q stationary condition

$$\nabla_q\mathcal{I}(r,q=\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}) = 0_d \quad\Rightarrow\quad r = r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}),$$

thus the claimed identity. Therefore

$$\begin{aligned}\tfrac{1}{n}I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big) &\geq \mathbb{E}_{\eta,\lambda}\textstyle\int_0^1 dt\,\sup_q\mathcal{I}\big(r^*(\mathbb{E}\langle Q\rangle_{t,\tilde{\epsilon}}),q\big) + o_n(1) \\ &\geq \inf\nolimits_{r\in\mathcal{S}_d^+}\sup\nolimits_{q\in\mathcal{S}_d^+}\mathcal{I}(r,q) + o_n(1).\end{aligned}$$

Finally, taking the $\liminf_{n\to+\infty}$ yields the converse bound of (10), and thus ends the proof of the mutual information replica symmetric variational formula. ■

IV. Asymptotic MMSE Matrix

In this section we prove the statement about the MMSE matrix in theorem II.1. Let f be given as in (2) and define $f_n : \mathcal{S}_d^+ \mapsto \mathbb{R}$ according to

$$f_n(S) \equiv \tfrac{1}{n}I\big(\boldsymbol{X};(\tilde{\boldsymbol{Y}},\boldsymbol{Y})\big).$$

By the matrix I-MMSE relation, f_n is differentiable on $\mathcal{S}_d^{++}$ with gradient

$$\nabla f_n(S) = \mathsf{MMSE}(\boldsymbol{X} \mid \tilde{\boldsymbol{Y}},\boldsymbol{Y}).$$

Meanwhile, f is concave because it is the minimum of a family of concave functions and thus it is differentiable almost everywhere. Similar to [31, Appendix A.3], we will show that pointwise convergence of f_n to f combined with the concavity of f_n implies convergence of the gradients everywhere f is differentiable.

To proceed, fix any point $S \in \mathcal{S}_d^{++}$ such that f is differentiable. By Griffiths' lemma [36, pg. 25] and the pointwise convergence of f_n to f, the directional derivates satisfy

$$\limsup_{n\to\infty}\mathrm{Tr}(T\nabla f_n(S)) = \mathrm{Tr}(T\nabla f(S))$$

for all $T \in \mathcal{S}_d$. Moreover both $\|\nabla f_n(S)\|$ and $\|\nabla f(S)\|$ are bounded in terms a constant C that depends only on the support of p_X. Consequently, pointwise convergence of the mapping $T \mapsto \mathrm{Tr}(T\nabla f_n(S))$ on $T \mapsto \mathrm{Tr}(T\nabla f(S))$ implies uniform convergence on any compact subset of $\mathcal{S}_d$, and we have

$$\begin{aligned}&\limsup_{n\to\infty}\|\nabla f_n(S) - \nabla f(S)\| \\ &\quad= \limsup_{n\to\infty}\sup_{T\in\mathcal{S}_d\,:\,\|T\|\leq 1}\mathrm{Tr}(T(\nabla f_n(S) - \nabla f(S)) = 0.\end{aligned}$$

We note that for points where f is not differentiable (i.e., the optimization in (2) does not have a unique solution), this argument can be adapted to provide lower and upper bounds on the MMSE matrix in terms of the subdifferential of f.

V. Conclusion and Perspectives

We characterized the information-theoretic limits of a class of multiview spiked matrix models. Our analysis both unifies and significantly extends the existing body of work. One important consequence of our results is to establish the tightness of the bounds obtained previously for community detection with correlated degree-balanced stochastic block models [31], [32].

The advances in this paper are made possible by a novel modification of the adaptive interpolation method [27], [39]. At a high level, this modification provides a decoupling between two main components of the method, namely the *interpolation path* and the *perturbation* (used for proving concentration of the overlap), and thus completely bypasses a number of non-trivial technical issues that previously limited the scope of the method. As a consequence, the method can be applied generically to a larger class of inference problems.

References

[1] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random cdma systems," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5590–5613, Nov 2010.

[2] J. Barbier, M. Dia, and N. Macris, "Proof of threshold saturation for spatially coupled sparse superposition codes," in *2016 IEEE Int. Symp. Inform. Theory*, July 2016, pp. 1173–1177.

[3] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse superposition codes via approximate message passing decoding," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1476–1500, 2017.

[4] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," in *IEEE Int. Symp. Inform. Theory*, 2016, pp. 665–669.

[5] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 625–632.

[6] J. Barbier, N. Macris, A. Maillard, and F. Krzakala, "The mutual information in random linear estimation beyond iid matrices," in *2018 IEEE Int. Symp. Inform. Theory*. IEEE, 2018, pp. 1390–1394.

[7] G. Reeves and H. D. Pfister, "Understanding phase transitions via mutual information and mmse," *arXiv preprint arXiv:1907.02095*, 2019.

[8] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5451–5460, 2019.

[9] M. Gabrié, A. Manoel, C. Luneau, N. Macris, F. Krzakala, L. Zdeborová *et al.*, "Entropy and mutual information in models of deep neural networks," in *Advances in Neural Information Processing Systems*, 2018, pp. 1821–1831.

[10] B. Aubin, A. Maillard, J. Barbier, F. Krzakala, N. Macris, and L. Zdeborová, "The committee machine: Computational to statistical gaps in learning a two-layers neural network," in *Advances in Neural Information Processing Systems*, 2018, pp. 3223–3234.

[11] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 04 2001.

[12] J. Baik, G. B. Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Annals of Probability*, p. 1643, 2005.

[13] A. A. Amini and M. J. Wainwright, "High-dimensional analysis of semidefinite relaxations for sparse principal components," *Ann. Statist.*, vol. 37, no. 5B, pp. 2877–2921, 10 2009.

[14] R. Krauthgamer, B. Nadler, and D. Vilenchik, "Do semidefinite relaxations solve sparse pca up to the information limit?" *Ann. Statist.*, vol. 43, no. 3, pp. 1300–1322, 06 2015.

[15] Y. Deshpande and A. Montanari, "Sparse pca via covariance thresholding," *Journal of Machine Learning Research*, vol. 17, no. 141, pp. 1–41, 2016.

[16] T. Ma and A. Wigderson, "Sum-of-squares lower bounds for sparse pca," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'15. Cambridge, MA, USA: MIT Press, 2015, pp. 1612–1620.

[17] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse pca," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 2197–2201.

[18] A. K. Fletcher and S. Rangan, "Iterative reconstruction of rank-one matrices in noise," *Information and Inference: A Journal of the IMA*, vol. 7, no. 3, pp. 531–562, 2018.

[19] S. S. Mannelli, G. Biroli, C. Cammarota, F. Krzakala, P. Urbani, and L. Zdeborová, "Marvels and pitfalls of the langevin algorithm in noisy high-dimensional inference," *arXiv preprint arXiv:1812.09066*, 2018.

[20] T. Lesieur, F. Krzakala, and L. Zdeborová, "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, no. 7, p. 073403, 2017.

[21] S. B. Korada and N. Macris, "Exact solution of the gauge symmetric p-spin glass model on a complete graph," *Journal of Statistical Physics*, vol. 136, no. 2, pp. 205–230, 2009.

[22] F. Krzakala, J. Xu, and L. Zdeborová, "Mutual information in rank-one matrix estimation," in *2016 IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 71–75.

[23] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Advances in Neural Information Processing Systems*, 2016, pp. 424–432.

[24] J. Barbier, M. Dia, N. Macris, F. Krzakala, and L. Zdeborová, "Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method," 2018.

[25] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *Probability Theory and Related Fields*, vol. 173, no. 3-4, pp. 859–929, 2019.

[26] J. Barbier, N. Macris, and L. Miolane, "The Layered Structure of Tensor Estimation and its Mutual Information," in *47th Annual Allerton Conference on Communication, Control, and Computing*, 2017.

[27] J. Barbier and N. Macris, "The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference," *Probability Theory and Related Fields*, Oct 2018.

[28] A. El Alaoui and F. Krzakala, "Estimation in the spiked wigner model: a short proof of the replica formula," in *2018 IEEE Int. Symp. Inform. Theory*. IEEE, 2018, pp. 1874–1878.

[29] J. Barbier and N. Macris, "0-1 phase transitions in sparse spiked matrix estimation," *arXiv preprint arXiv:1911.05030*, 2019.

[30] Y. Deshpande, E. Abbe, and A. Montanari, "Asymptotic mutual information for the two-groups stochastic block model," *arXiv:1507.08685*, 2015.

[31] G. Reeves, V. Mayya, and A. Volfovsky, "The geometry of community detection via the mmse matrix," in *2019 IEEE Int. Symp. Inform. Theory*. IEEE, 2019, pp. 400–404.

[32] V. Mayya and G. Reeves, "Mutual information in community detection with covariate information and correlated networks," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 602–607.

[33] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov 2009.

[34] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb 2011.

[35] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[36] M. Talagrand, *Mean field models for spin glasses: Volume I: Basic examples*. Springer Science & Business Media, 2010, vol. 54.

[37] F. Guerra and F. L. Toninelli, "The thermodynamic limit in mean field spin glass models," *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, 2002.

[38] F. Guerra, "Replica broken bounds in the mean field spin glass model," *Comm. Math. Phys.*, vol. 233, pp. 1–12, 2003.

[39] J. Barbier and N. 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*, vol. 52, no. 29, p. 294002, jun 2019.

[40] J.-C. Mourrat, "Hamilton-Jacobi equations for mean-field disordered systems," 2018, [Online]. Available https://arxiv.org/abs/1811.01432.

[41] J. Barbier, C. Luneau, and N. Macris, "Mutual information for low-rank even-order symmetric tensor factorization," *arXiv preprint arXiv:1904.04565*, 2019.

[42] L. Le Cam, *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

[43] M. Payaró, M. Gregori, and D. Palomar, "Yet another entropy power inequality with an application," in *2011 International Conference on Wireless Communications and Signal Processing (WCSP)*.

[44] G. Reeves, H. D. Pfister, and A. Dytso, "Mutual information as a function of matrix snr for linear gaussian channels," in *2018 IEEE Int. Symp. Inform. Theory*. IEEE, 2018, pp. 1754–1758.

[45] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Wiley, 2007.

[46] J. Barbier, "Overlap matrix concentration in optimal Bayesian inference," *Information and Inference: A Journal of the IMA*, 05 2020, iaaa008. [Online]. Available: https://doi.org/10.1093/imaiai/iaaa008

[47] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE transactions on information theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.

[48] A. Lasota, A. Strauss, and W. Walter, "Infinite systems of differential inequalities defined recursively," *Journal of Differential Equations*, vol. 9, no. 1, pp. 93–107, 1970.

[49] G. Reeves, "Information-theoretic limits for the matrix tensor product," 2020, preprint.