Proceedings of Machine Learning Research vol 65:1–5, 2017

Fundamental limits of symmetric low-rank matrix estimation

Marc Lelarge MARC.LELARGE@ENS.FR
Safran Tech, Magny-Les-Hameaux, France
Département d'informatique, École normale supérieure, CNRS, PSL Research University, Paris, France
INRIA
Léo Miolane LEO.MIOLANE@INRIA.FR
Département d'informatique, École normale supérieure, CNRS, PSL Research University, Paris, France
INRIA

Abstract

We consider the high-dimensional inference problem where the signal is a low-rank symmetric matrix which is corrupted by an additive Gaussian noise. Given a probabilistic model for the low-rank matrix, we compute the limit in the large dimension setting for the mutual information between the signal and the observations, as well as the matrix minimum mean square error, while the rank of the signal remains constant. We unify and generalize a number of recent works on PCA, sparse PCA, submatrix localization or community detection by computing the information-theoretic limits for these problems in the high noise regime. This allows to locate precisely the information-theoretic thresholds for the above mentioned problems.[1]

Keywords: Matrix factorization, PCA, community detection, spin glasses

1. Low-rank matrix estimation

The estimation of a low-rank matrix observed through a noisy channel is a fundamental problem in statistical inference with applications in machine learning, signal processing or information theory. We shall consider the high dimensional setting where the low-rank matrix to estimate is symmetric and where the noise is additive and Gaussian.

Let P_0 be a probability distribution over $\mathbb{R}^k$ ($k \in \mathbb{N}^*$ is fixed) with finite second moment. Let $\mathbf{X}_i \overset{\text{i.i.d.}}{\sim} P_0$ and suppose that we observe for $1 \leq i < j \leq n$:

$$Y_{i,j} = \sqrt{\frac{\lambda}{n}} \mathbf{X}_i^\intercal \mathbf{X}_j + Z_{i,j} \tag{1}$$

where $Z_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. The goal here is to recover the low-rank signal $\mathbf{X}\mathbf{X}^\intercal$ from the observation $\mathbf{Y}$. Notice that we suppose here to observe only the coefficients of $\sqrt{\lambda/n}\mathbf{X}\mathbf{X}^\intercal + \mathbf{Z}$ that are above the diagonal. The case where all the coefficients are observed can be directly deduced from this case.

Depending on the choice of the prior P_0, (1) can encode many classical statistical problems such as PCA, sparse PCA, submatrix localization or community detection. These examples are detailed in the full version of this work. Define the minimal mean square error (MMSE) for this statistical

1. Extended abstract. Full version appears as [arXiv:1611.03888v3]

problem:

$$\mathrm{MMSE}_n(\lambda) = \min_{\hat{\theta}} \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbb{E}\left[\left(\mathbf{X}_i^\intercal \mathbf{X}_j - \hat{\theta}_{i,j}(\mathbf{Y})\right)^2\right]$$

where the minimum is taken over all estimators $\hat{\theta}$ (i.e. measurable functions of the observations $\mathbf{Y}$). We aim at computing the limit of MMSE_n and the mutual information $\frac{1}{n}I(\mathbf{X};\mathbf{Y})$ as n goes to infinity.

2. The replica symmetric formula

We define the function

$$\mathcal{F} : (\lambda, \mathbf{q}) \in \mathbb{R} \times S_k^+ \mapsto \mathbb{E} \log \int dP_0(\mathbf{x}) \exp\left(\sqrt{\lambda}(\mathbf{Z}^\intercal \mathbf{q}^{1/2}\mathbf{x}) + \lambda \mathbf{x}^\intercal \mathbf{q} \mathbf{X} - \frac{\lambda}{2}\mathbf{x}^\intercal \mathbf{q} \mathbf{x}\right) - \frac{\lambda}{4}\|\mathbf{q}\|^2$$

where S_k^+ denote the set of $k \times k$ symmetric positive-semidefinite matrices. $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_k)$ and $\mathbf{X} \sim P_0$ are independent random variables. The limits of the MMSE and the mutual information has been conjectured in Lesieur et al. (2015b) according to powerful, but non-rigorous, statistical physics arguments. They are given by the following "replica symmetric" formula.

Theorem 1 *For $\lambda > 0$, we have*

$$\lim_{n \to +\infty} \frac{1}{n} I(\mathbf{X};\mathbf{Y}) = \frac{\lambda \|\mathbb{E}_{P_0} \mathbf{X}\mathbf{X}^\intercal\|^2}{4} - \sup_{\mathbf{q} \in S_k^+} \mathcal{F}(\lambda, \mathbf{q}),$$

For almost all $\lambda > 0$, all the maximizers $\mathbf{q}$ of $\mathbf{q} \in S_k^+ \mapsto \mathcal{F}(\lambda, \mathbf{q})$ have the same norm $\|\mathbf{q}\| = q^*(\lambda)$ and

$$\mathrm{MMSE}_n(\lambda) \xrightarrow[n \to \infty]{} \|\mathbb{E}_{P_0} \mathbf{X}\mathbf{X}^\intercal\|^2 - q^*(\lambda)^2$$

In the rank-one case ($k = 1$), Barbier et al. (2016) proved Theorem 1 for discrete P_0 but under the restrictive assumption that the function $q \mapsto \mathcal{F}(\lambda, q)$ is required to have at most three stationary points.

Notice that the "dummy estimator" $\hat{\theta}_{i,j}(\mathbf{Y}) = \mathbb{E}[\mathbf{X}_i \mathbf{X}_j^\intercal]$ achieves has a mean square error equal to DMSE $= \|\mathbb{E}\mathbf{X}\mathbf{X}^\intercal\|^2 - \|(\mathbb{E}\mathbf{X})(\mathbb{E}\mathbf{X})^\intercal\|^2$. We deduce from Theorem 1 that

$$\lambda_c := \inf \left\{ \lambda > 0 \mid q^*(\lambda) > \|(\mathbb{E}\mathbf{X})(\mathbb{E}\mathbf{X})^\intercal\| \right\}$$

is the information-theoretic threshold for the estimation problem (1). Namely,

- if $\lambda > \lambda_c$, then $\lim_{n \to \infty} \mathrm{MMSE}_n <$ DMSE: one can estimate $\mathbf{X}\mathbf{X}^\intercal$ better than chance as n goes to infinity.
- if $\lambda < \lambda_c$, then $\lim_{n \to \infty} \mathrm{MMSE}_n =$ DMSE: it is impossible to retrieve asymptotically $\mathbf{X}\mathbf{X}^\intercal$ better than a "random guess".

References

IEEE International Symposium on Information Theory, ISIT 2015, Hong Kong, China, June 14-19, 2015, 2015. IEEE. ISBN 978-1-4673-7704-1. URL http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7270906.

Emmanuel Abbe and Colin Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint arXiv:1512.09080*, 2015.

Michael Aizenman, Robert Sims, and Shannon L Starr. Extended variational principle for the sherrington-kirkpatrick spin-glass model. *Physical Review B*, 68(21):214403, 2003.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.

Afonso S Bandeira, Nicolas Boumal, and Amit Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, pages 1–23, 2016.

Jess Banks, Cristopher Moore, Nicolas Verzelen, Roman Vershynin, and Jiaming Xu. Information-theoretic bounds and phase transitions in clustering, sparse pca, and submatrix localization. *arXiv preprint arXiv:1607.05222v2*, 2016.

Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems*, pages 424–432, 2016.

Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Francesco Caltagirone, Marc Lelarge, and Léo Miolane. Recovering asymmetric communities in the stochastic block model. *arXiv preprint arXiv:1610.03680*, 2016.

Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborova. Information-theoretic thresholds from the cavity method. *arXiv preprint arXiv:1611.00814*, 2016.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

Yash Deshpande and Emmanuel Abbe. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference*, page iaw017, 2016.

Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.

Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics*, 272(1):185–228, 2007.

Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005.

Bruce Hajek, Yihong Wu, and Jiaming Xu. Information limits for recovering a hidden community. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 1894–1898. IEEE, 2016.

Yukito Iba. The nishimori line and bayesian statistics. *Journal of Physics A: Mathematical and General*, 32(21):3875, 1999.

Satish Babu Korada and Nicolas Macris. Exact solution of the gauge symmetric p-spin glass model on a complete graph. *Journal of Statistical Physics*, 136(2):205–230, 2009.

Satish Babu Korada and Nicolas Macris. Tight bounds on the capacity of binary input random cdma systems. *IEEE Transactions on Information Theory*, 56(11):5590–5613, 2010.

Satish Babu Korada and Andrea Montanari. Applications of the lindeberg principle in communications and statistical learning. *IEEE Transactions on Information Theory*, 57(4):2440–2450, 2011.

Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *Information Theory Workshop (ITW), 2016 IEEE*, pages 71–75. IEEE, 2016.

Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *53rd Annual Allerton Conference on Communication, Control, and Computing, Allerton 2015, Allerton Park & Retreat Center, Monticello, IL, USA, September 29 - October 2, 2015*, pages 680–687. IEEE, 2015a. ISBN 978-1-5090-1824-6. doi: 10.1109/ALLERTON.2015.7447070. URL http://dx.doi.org/10.1109/ALLERTON.2015.7447070.

Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse PCA. In *IEEE International Symposium on Information Theory, ISIT 2015, Hong Kong, China, June 14-19, 2015* DBL (2015), pages 1635–1639. ISBN 978-1-4673-7704-1. doi: 10.1109/ISIT.2015.7282733. URL http://dx.doi.org/10.1109/ISIT.2015.7282733.

Cyril Méasson, Andrea Montanari, Thomas J Richardson, and Rüdiger Urbanke. The generalized area theorem and some of its consequences. *IEEE Transactions on Information Theory*, 55(11):4793–4821, 2009.

Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Co Inc, 1987.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

Andrea Montanari. Estimating random variables from random sparse observations. *European Transactions on Telecommunications*, 19(4):385–403, 2008.

Andrea Montanari. Finding one community in a sparse graph. *Journal of Statistical Physics*, 161 (2):273–299, 2015.

Joe Neeman and Praneeth Netrapalli. Non-reconstructability in the stochastic block model. *arXiv preprint arXiv:1404.6304*, 2014.

Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*, volume 111. Clarendon Press, 2001.

Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.

Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca for spiked random matrices and synchronization. *arXiv preprint arXiv:1609.05573*, 2016.

Sundeep Rangan and Alyson K Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1246–1250. IEEE, 2012.

Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Spectral detection in the censored block model. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1184–1188. IEEE, 2015.

Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Clustering from sparse pairwise measurements. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 780–784. IEEE, 2016.

Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

Michel Talagrand. *Mean field models for spin glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*, volume 55. Springer Science & Business Media, 2011.

Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.