

# **Spectral Methods for Data Science: A Statistical Perspective**

**Other titles in Foundations and Trends® in Machine Learning**

*Tensor Regression*

Jiani Liu, Ce Zhu, Zhen Long and Yipeng Liu

ISBN: 978-1-68083-886-2

*Minimum-Distortion Embedding*

Akshay Agrawal, Alnur Ali and Stephen Boyd

ISBN: 978-1-68083-888-6

*Graph Kernels: State-of-the-Art and Future Challenges*

Karsten Borgwardt, Elisabetta Ghisu, Felipe Linares-López, Leslie O'Bray and Bastian Rieck

ISBN: 978-1-68083-770-4

*Data Analytics on Graphs Part III: Machine Learning on Graphs, from Graph Topology to Applications*

Ljubiša Stanković, Danilo Mandić, Miloš Daković, Miloš Brajović, Bruno Scalzo, Shengxi Li and Anthony G. Constantinides

ISBN: 978-1-68083-982-16

*Data Analytics on Graphs Part II: Signals on Graphs*

Ljubiša Stanković, Danilo Mandić, Miloš Daković, Miloš Brajović, Bruno Scalzo, Shengxi Li and Anthony G. Constantinides

ISBN: 978-1-68083-982-1

*Data Analytics on Graphs Part I: Graphs and Spectra on Graphs*

Ljubiša Stanković, Danilo Mandić, Miloš Daković, Miloš Brajović, Bruno Scalzo, Shengxi Li and Anthony G. Constantinides

ISBN: 978-1-68083-982-1

# Spectral Methods for Data Science: A Statistical Perspective

---

**Yuxin Chen**

Princeton University  
yuxin.chen@princeton.edu

**Yuejie Chi**

Carnegie Mellon University  
yuejiechi@cmu.edu

**Jianqing Fan**

Princeton University  
jqfan@princeton.edu

**Cong Ma**

University of Chicago  
congma@uchicago.edu

**now**

the essence of knowledge

Boston — Delft

## Foundations and Trends<sup>®</sup> in Machine Learning

*Published, sold and distributed by:*

now Publishers Inc.  
PO Box 1024  
Hanover, MA 02339  
United States  
Tel. +1-781-985-4510  
[www.nowpublishers.com](http://www.nowpublishers.com)  
[sales@nowpublishers.com](mailto:sales@nowpublishers.com)

*Outside North America:*

now Publishers Inc.  
PO Box 179  
2600 AD Delft  
The Netherlands  
Tel. +31-6-51115274

The preferred citation for this publication is

Y. Chen, Y. Chi, J. Fan and C. Ma. *Spectral Methods for Data Science: A Statistical Perspective*. Foundations and Trends<sup>®</sup> in Machine Learning, vol. 14, no. 5, pp. 566–806, 2021.

ISBN: 978-1-68083-897-8

© 2021 Y. Chen, Y. Chi, J. Fan and C. Ma

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: [www.copyright.com](http://www.copyright.com)

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; [www.nowpublishers.com](http://www.nowpublishers.com); [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, [www.nowpublishers.com](http://www.nowpublishers.com); e-mail: [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

# Foundations and Trends<sup>®</sup> in Machine Learning

Volume 14, Issue 5, 2021

## Editorial Board

### Editor-in-Chief

**Michael Jordan**

University of California, Berkeley  
United States

### Editors

Peter Bartlett  
*UC Berkeley*

Yoshua Bengio  
*Université de Montréal*

Avrim Blum  
*Toyota Technological  
Institute*

Craig Boutilier  
*University of Toronto*

Stephen Boyd  
*Stanford University*

Carla Brodley  
*Northeastern University*

Inderjit Dhillon  
*Texas at Austin*

Jerome Friedman  
*Stanford University*

Kenji Fukumizu  
*ISM*

Zoubin Ghahramani  
*Cambridge University*

David Heckerman  
*Amazon*

Tom Heskes  
*Radboud University*

Geoffrey Hinton  
*University of Toronto*

Aapo Hyvarinen  
*Helsinki IIT*

Leslie Pack Kaelbling  
*MIT*

Michael Kearns  
*UPenn*

Daphne Koller  
*Stanford University*

John Lafferty  
*Yale*

Michael Littman  
*Brown University*

Gabor Lugosi  
*Pompeu Fabra*

David Madigan  
*Columbia University*

Pascal Massart  
*Université de Paris-Sud*

Andrew McCallum  
*University of  
Massachusetts Amherst*

Marina Meila  
*University of Washington*

Andrew Moore  
*CMU*

John Platt  
*Microsoft Research*

Luc de Raedt  
*KU Leuven*

Christian Robert  
*Paris-Dauphine*

Sunita Sarawagi  
*IIT Bombay*

Robert Schapire  
*Microsoft Research*

Bernhard Schoelkopf  
*Max Planck Institute*

Richard Sutton  
*University of Alberta*

Larry Wasserman  
*CMU*

Bin Yu  
*UC Berkeley*

## Editorial Scope

### Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

### Information for Librarians

Foundations and Trends® in Machine Learning, 2021, Volume 14, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Motivating applications . . . . .	4
1.2	A modern statistical perspective . . . . .	11
1.3	Organization . . . . .	12
1.4	What is not here and complementary readings . . . . .	13
1.5	Notation . . . . .	14
<b>2</b>	<b>Classical spectral analysis: <math>l_2</math> perturbation theory</b>	<b>16</b>
2.1	Preliminaries: Basics of matrix analysis . . . . .	17
2.2	Preliminaries: Distance and angles between subspaces . . . . .	18
2.3	Perturbation theory for eigenspaces . . . . .	22
2.4	Perturbation theory for singular subspaces . . . . .	27
2.5	Eigenvector perturbation for probability transition matrices . . . . .	30
2.6	Appendix: Proofs of auxiliary lemmas in Section 2.2 . . . . .	34
2.7	Notes . . . . .	39
<b>3</b>	<b>Applications of <math>l_2</math> perturbation theory to data science</b>	<b>41</b>
3.1	Preliminaries: Matrix tail bounds . . . . .	41
3.2	Low-rank matrix denoising . . . . .	45
3.3	Principal component analysis and factor models . . . . .	48
3.4	Graph clustering and community recovery . . . . .	55
3.5	Clustering in Gaussian mixture models . . . . .	61

3.6	Ranking from pairwise comparisons . . . . .	77
3.7	Phase retrieval and solving quadratic systems of equations . . . . .	86
3.8	Matrix completion . . . . .	98
3.9	Tensor completion . . . . .	106
3.10	Notes . . . . .	118
<b>4</b>	<b>Fine-grained analysis: <math>l_\infty</math> and <math>l_{2,\infty}</math> perturbation theory</b>	<b>128</b>
4.1	Leave-one-out analysis: An illustrative example . . . . .	129
4.2	$l_{2,\infty}$ eigenspace perturbation under independent noise . . . . .	136
4.3	$l_{2,\infty}$ singular subspace perturbation under independent noise . . . . .	141
4.4	Application: Entrywise guarantees for matrix completion . . . . .	143
4.5	Application: Exact community recovery . . . . .	144
4.6	Distributional theory and uncertainty quantification . . . . .	153
4.7	Application: Confidence intervals for matrix completion . . . . .	158
4.8	Appendix A: Proof of Theorem 4.2 . . . . .	164
4.9	Appendix B: Proof of Corollary 4.3 . . . . .	178
4.10	Appendix C: Proof of Theorem 4.4 . . . . .	182
4.11	Appendix D: Proof of Theorem 4.10 . . . . .	185
4.12	Appendix E: Proof of Theorem 4.11 . . . . .	192
4.13	Notes . . . . .	198
<b>5</b>	<b>Concluding remarks and open problems</b>	<b>201</b>
	<b>Acknowledgements</b>	<b>204</b>
	<b>References</b>	<b>206</b>



# Spectral Methods for Data Science: A Statistical Perspective

Yuxin Chen<sup>1</sup>, Yuejie Chi<sup>2</sup>, Jianqing Fan<sup>3</sup> and Cong Ma<sup>4</sup>

<sup>1</sup>*Princeton University; [yuxin.chen@princeton.edu](mailto:yuxin.chen@princeton.edu)*

<sup>2</sup>*Carnegie Mellon University; [yuejiechi@cmu.edu](mailto:yuejiechi@cmu.edu)*

<sup>3</sup>*Princeton University; [jqfan@princeton.edu](mailto:jqfan@princeton.edu)*

<sup>4</sup>*University of Chicago; [congma@uchicago.edu](mailto:congma@uchicago.edu)*

---

## ABSTRACT

Spectral methods have emerged as a simple yet surprisingly effective approach for extracting information from massive, noisy and incomplete data. In a nutshell, spectral methods refer to a collection of algorithms built upon the eigenvalues (resp. singular values) and eigenvectors (resp. singular vectors) of some properly designed matrices constructed from data. A diverse array of applications have been found in machine learning, imaging science, financial and econometric modeling, and signal processing, including recommendation systems, community detection, ranking, structured matrix recovery, tensor data estimation, joint shape matching, blind deconvolution, financial investments, risk managements, treatment evaluations, causal inference, amongst others. Due to their simplicity and effectiveness, spectral methods are not only used as a stand-alone estimator, but also frequently employed to facilitate other more sophisticated algorithms to enhance performance.

While the studies of spectral methods can be traced back to classical matrix perturbation theory and the method of

moments, the past decade has witnessed tremendous theoretical advances in demystifying their efficacy through the lens of statistical modeling, with the aid of concentration inequalities and non-asymptotic random matrix theory. This monograph aims to present a systematic, comprehensive, yet accessible introduction to spectral methods from a modern statistical perspective, highlighting their algorithmic implications in diverse large-scale applications. In particular, our exposition gravitates around several central questions that span various applications: how to characterize the sample efficiency of spectral methods in reaching a target level of statistical accuracy, and how to assess their stability in the face of random noise, missing data, and adversarial corruptions? In addition to conventional  $\ell_2$  perturbation analysis, we present a systematic  $\ell_\infty$  and  $\ell_{2,\infty}$  perturbation theory for eigenspace and singular subspaces, which has only recently become available owing to a powerful “leave-one-out” analysis framework.

---

# 1

---

## Introduction

---

In contemporary science and engineering applications, the volume of available data is growing at an enormous rate. The emergence of this trend is due to recent technological advances that have enabled the collection, transmission, storage and processing of data from every corner of our life, in the forms of images, videos, network traffic, email logs, electronic health records, genomic and genetic measurements, high-frequency financial trades, grocery transactions, online exchanges, and so on. In the meantime, modern applications often require reasonings about an unprecedented scale of features or parameters of interest. This gives rise to the pressing demand of developing *low-complexity* algorithms that can effectively distill actionable insights from large-scale and high-dimensional data. In addition to the curse of dimensionality, the challenge is further compounded when the data in hand are noisy, messy, and contain missing features.

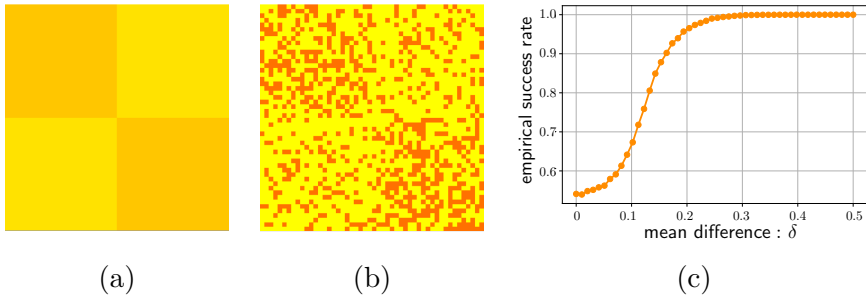
Towards addressing the above challenges, *spectral methods* have emerged as a simple yet surprisingly effective approach to information extraction from massive and noisy data. In a nutshell, spectral methods refer to a collection of algorithms built upon the eigenvectors (resp. singular vectors) and eigenvalues (resp. singular values) of some

properly designed matrices generated from data. Remarkably, spectral methods lend themselves to a diverse array of applications in practice, including community detection in networks (Newman, 2006; Abbe, 2017; Rohe *et al.*, 2011; McSherry, 2001), angular synchronization in cryo-EM (Singer and Shkolnisky, 2011; Singer, 2011), joint image alignment (Chen and Candès, 2018), clustering (Von Luxburg, 2007; Ng *et al.*, 2002), ranking (Negahban *et al.*, 2016; Chen and Suh, 2015; Chen *et al.*, 2019b), dimensionality reduction (Belkin and Niyogi, 2003), low-rank matrix estimation (Achlioptas and McSherry, 2007; Keshavan *et al.*, 2010), tensor estimation (Montanari and Sun, 2018; Cai *et al.*, 2019a), covariance and precision matrix estimation (Fan *et al.*, 2013; Fan *et al.*, 2021b), shape reconstruction (Li and Hero, 2004), econometric and financial modeling (Fan *et al.*, 2021a), among others. Motivated by their applicability to numerous real-world problems, this monograph seeks to offer a unified and comprehensive treatment towards establishing the theoretical underpinnings for spectral methods, particularly through a statistical lens.

## 1.1 Motivating applications

At the heart of spectral methods is the idea that the eigenvectors or singular vectors of certain data matrices reveal crucial information pertaining to the targets of interest. We single out a few examples that epitomize this idea.

**Clustering.** Clustering corresponds to the grouping of individuals based on their mutual similarities, which constitutes a fundamental task in unsupervised learning and spans numerous applications such as image segmentation (e.g., grouping pixels based on the objects they represent in an image) (Browet *et al.*, 2011) and community detection (e.g., grouping users on the basis of their social circles) (Fortunato and Hric, 2016). For concreteness, let us take a look at a simple scenario with  $n$  individuals such that: (1) there exists a latent partitioning that divides all individuals into two groups, with the first  $n/2$  individuals belonging to the first group and the rest belonging to the second group (without loss of generality); and (2) we observe pairwise similarity



**Figure 1.1:** Spectral methods for clustering. We plot in (a) an ideal structure of the adjacency matrix  $\mathbf{A}$  in (1.1), and in (b) a noisy version which is a realization from the stochastic block model, where  $A_{i,j}$  is an independent Bernoulli variable with mean  $\frac{1+\delta}{2}$  (resp.  $\frac{1-\delta}{2}$ ) if  $i$  and  $j$  belong to the same group (resp. different groups). We report in (c) the empirical success rate of the spectral method over 200 Monte Carlo trials in correctly clustering  $n = 100$  individuals as the mean difference  $\delta$  varies.

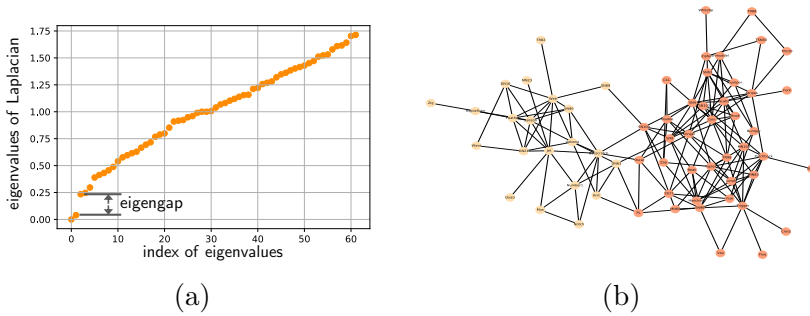
measurements generated based on their group memberships. Ideally, if we know whether any two individuals belong to the same group or not, then we can form an adjacency matrix  $\mathbf{A} = [A_{i,j}]_{1 \leq i,j \leq n}$  such that

$$A_{i,j} = \begin{cases} 1, & \text{if } (i,j) \text{ belongs to the same group,} \\ 0, & \text{else.} \end{cases} \quad (1.1)$$

As a key observation, this matrix  $\mathbf{A}$ , as illustrated in Figure 1.1(a), turns out to be a rank-2 matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_{n/2} \mathbf{1}_{n/2}^\top & \\ & \mathbf{1}_{n/2} \mathbf{1}_{n/2}^\top \end{bmatrix} = \frac{1}{2} \mathbf{1}_n \mathbf{1}_n^\top + \frac{1}{2} \begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix} \begin{bmatrix} \mathbf{1}_{n/2}^\top & -\mathbf{1}_{n/2}^\top \end{bmatrix},$$

where  $\mathbf{1}_n$  represents an  $n$ -dimensional all-one vector. After subtracting  $\frac{1}{2} \mathbf{1}_n \mathbf{1}_n^\top$  from  $\mathbf{A}$ , the eigenvector  $\mathbf{u}_2 := \begin{bmatrix} \mathbf{1}_{n/2}^\top & -\mathbf{1}_{n/2}^\top \end{bmatrix}$  of the remaining component uncovers the underlying group structure; namely, all positive entries of  $\mathbf{u}_2$  represent one group, with all negative entries of  $\mathbf{u}_2$  reflecting another group. In reality, however, we typically only get to collect imprecise information about whether two individuals belong to the same group, thus resulting in a corrupted version of  $\mathbf{A}$  (see Figure 1.1(b)). Fortunately, the eigenvector (the one corresponding to  $\mathbf{u}_2$  above) of the observed data matrix (with proper arrangement) might



**Figure 1.2:** Illustration of spectral clustering for 62 dolphins residing in Doubtful Sound, New Zealand. (a) plots the spectrum of the Laplacian matrix of an undirected social network of frequent associations, and (b) illustrates the two communities recovered using the penultimate eigenvector of the Laplacian matrix. Data source: Lusseau *et al.* (2003).

continue to be informative, as long as the noise level is not overly high. To illustrate the practical applicability, we plot in Figure 1.1(c) the numerical performance of this approach, which allows for perfect clustering of all individuals for a wide range of noisy scenarios. Similar ideas continue to fare well on the clustering of real data, where we illustrate in Figure 1.2 that the penultimate eigenvector of a Laplacian matrix (also known as the Fiedler vector) of an undirected social network reveals two communities of 62 dolphins residing in Doubtful Sound, New Zealand.

**Principal component analysis (PCA).** PCA is arguably one of the most commonly employed tools for data exploration and visualization. Given a collection of data samples  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ , PCA seeks to identify a rank- $r$  subspace that explains most of the variability of the data. This is particularly well-grounded when, say, the sample vectors  $\{\mathbf{x}_i\}_{1 \leq i \leq n}$  reside primarily within a common rank- $r$  subspace—denoted by  $\mathbf{U}^*$ . To extract out this principal subspace, it is instrumental to examine the following sample covariance matrix

$$\mathbf{M} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top.$$

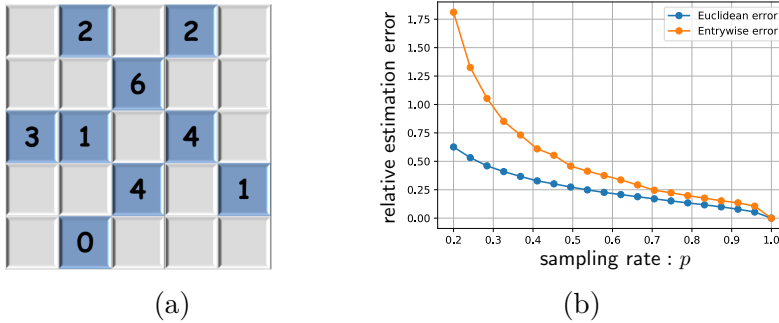
If all sample vectors approximately lie within  $\mathbf{U}^*$ , then one might be able to infer  $\mathbf{U}^*$  by inspecting the rank- $r$  leading eigenspace of  $\mathbf{M}$  (or its variants), provided that the signal-to-noise ratio exceeds some reasonable level. This reflects the role of spectral methods in enabling meaningful dimensionality reduction and factor analysis.

In practice, a key benefit of PCA is its ability to remove nuance factors in, and extract out salient features from, each data point. As an illustration, the first four images of Figure 1.3 are representative ones sampled from a face dataset (Georghiades *et al.*, 2001), which correspond to faces of the same person under different illumination and occlusion conditions. In contrast, the “eigenface” (Turk and Pentland, 1991) depicted in the last image of Figure 1.3 corresponds to the first principal component (i.e.,  $r = 1$ ), which effectively removes the nuance factors and highlights the feature of the face.



**Figure 1.3:** Illustration of the eigenface using the Cropped YaleB dataset (Georghiades *et al.*, 2001). The first four images are sampled from this dataset, representing typical images taken under different illumination conditions with various occlusions. The last one represents the eigenface (i.e., the first principal component) of this dataset.

**Matrix recovery in the face of missing data.** A proliferation of big-data applications has to deal with matrix estimation in the presence of missing data, either due to the infeasibility to acquire complete observations of a massive data matrix (Davenport and Romberg, 2016) such as the Netflix problem in recommender systems (as users only watch and rate a small fraction of movies), or because of the incentive to accelerate computation by means of sub-sampling (Mahoney, 2016). Imagine that we are asked to estimate a large matrix  $\mathbf{M}^* = [M_{i,j}^*]_{1 \leq i,j \leq n}$ , even though a dominant fraction of its entries are unseen. While in general we cannot predict anything about the missing entries, reliable



**Figure 1.4:** Spectral methods for matrix recovery with missing data, where (a) is an illustration of missing data and (b) reports the empirical estimation errors of spectral methods as the sampling rate  $p$  varies. Both the relative Euclidean error  $\frac{\|\widehat{\mathbf{M}} - \mathbf{M}^*\|_{\text{F}}}{\|\mathbf{M}^*\|_{\text{F}}}$  and the relative entrywise error  $\frac{\|\widehat{\mathbf{M}} - \mathbf{M}^*\|_{\infty}}{\|\mathbf{M}^*\|_{\infty}}$  are plotted (with  $\widehat{\mathbf{M}}$  denoting the matrix estimate and  $\|\cdot\|_{\infty}$  the entrywise  $\ell_{\infty}$  norm).

estimation might become possible if  $\mathbf{M}^*$  is known *a priori* to enjoy a low-rank structure, as is the case in many applications like structure from motion (Tomasi and Kanade, 1992) and sensor network localization (Javanmard and Montanari, 2013). This low-rank assumption motivates the use of spectral methods. More specifically, suppose the entries of  $\mathbf{M}^*$  are randomly sampled such that each entry is observed independently with probability  $p \in (0, 1]$ . An unbiased estimate  $\mathbf{M} = [M_{i,j}]_{1 \leq i, j \leq n}$  of  $\mathbf{M}^*$  can be readily obtained via rescaling and zero filling (also called the inverse probability weighting method):

$$M_{i,j} = \begin{cases} \frac{1}{p} M_{i,j}^*, & \text{if the } (i, j)\text{-th entry is observed,} \\ 0, & \text{else.} \end{cases}$$

To capture the assumed low-rank structure of  $\mathbf{M}^*$ , it is natural to resort to the best rank- $r$  approximation of  $\mathbf{M}$  (with  $r$  the true rank of  $\mathbf{M}^*$ ), computable through the rank- $r$  singular value decomposition of  $\mathbf{M}$ . Given its (trivial) success when  $p = 1$ , we expect the algorithm to perform well when  $p$  is close to 1. The key question, however, is where the algorithm stands if the vast majority of the entries is missing. While we shall illuminate this in Chapters 3 and 4, Figure 1.4 provides some immediate numerical assessment, which demonstrates the appealing



performance of spectral methods—in terms of both Euclidean and entrywise estimation errors—even when the missing rate is quite high.

**Ranking from pairwise comparisons.** Another important application of spectral methods arises from the context of ranking, a task of central importance in, say, web search and recommendation systems. In a variety of scenarios, humans find it difficult to simultaneously rank many items, but relatively easier to express pairwise preferences. This gives rise to the problem of ranking based on pairwise comparisons. More specifically, imagine we are given a collection of  $n$  items, and wish to identify top-ranked items based on pairwise preferences (with uncertainties in comparison outcomes) between observed pairs of items. A classical statistical model proposed by Bradley and Terry (1952) and Luce (2012) postulates the existence of a set of latent positive scores  $\{w_i^*\}_{1 \leq i \leq n}$ —each associated with an item—that determines the ranks of these items. The outcome of the comparison between items  $i$  and  $j$  is generated in a way that

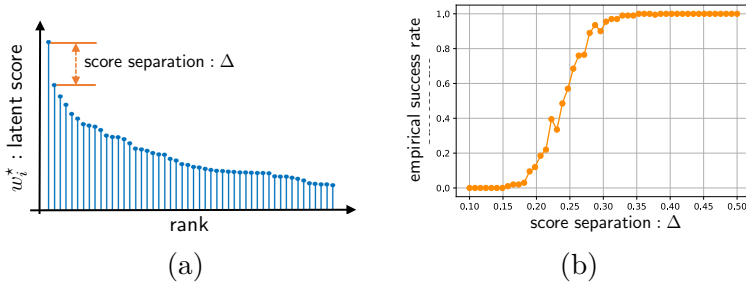
$$\mathbb{P}(i \text{ beats } j) = \frac{w_i^*}{w_i^* + w_j^*}, \quad 1 \leq i, j \leq n.$$

As it turns out, the preference scores are closely related to the stationary distribution of a Markov chain associated with the above probability kernel, thus forming the basis of spectral ranking algorithms. To elucidate it in a little more detail, let us construct a probability transition matrix  $\mathbf{P}^* = [P_{i,j}^*]_{1 \leq i, j \leq n}$  with

$$P_{i,j}^* = \begin{cases} \frac{1}{n} \cdot \frac{w_j^*}{w_i^* + w_j^*}, & \text{if } i \neq j, \\ 1 - \sum_{l:l \neq i} P_{i,l}^*, & \text{if } i = j. \end{cases}$$

Clearly, it forms a probability transition matrix since each element is non-negative and the entries in each row add up to one. It is straightforward to verify that the score vector  $\mathbf{w}^* := [w_i^*]_{1 \leq i \leq n}$  satisfies  $\mathbf{w}^{*\top} = \mathbf{w}^{*\top} \mathbf{P}^*$ , namely  $\mathbf{w}^*$  is a left eigenvector of  $\mathbf{P}^*$  associated with eigenvalue one. A candidate method then consists of (i) forming an unbiased estimate of  $\mathbf{P}^*$  (which can be easily obtained using pairwise comparison outcomes), (ii) computing its left eigenvector (in fact, the leading left eigenvector),

and (iii) reporting the ranking result in accordance with the order of the elements in this eigenvector. This spectral ranking scheme, which shares similar spirit with the celebrated *PageRank* algorithm (Page *et al.*, 1999), exhibits intriguing performance when identifying the top-ranked items, as showcased in the numerical experiments in Figure 1.5(b).



**Figure 1.5:** Spectral methods for ranking from pairwise comparisons. (a) illustrates the latent preference scores  $\{w_i^*\}$  that govern the ranking of items. The empirical success rates in correctly identifying the top-ranked item are plotted in (b) as  $\Delta$  varies, where  $\Delta$  represents the separation between the score of the top item and that of the second-ranked item.

**A unified theme.** In all preceding applications, the core ideas underlying the development of spectral methods can be described in a unified fashion:

1. Identify a key matrix  $\mathbf{M}^*$ —which is typically unobserved—whose eigenvectors or singular vectors disclose the information being sought after;
2. Construct a surrogate matrix  $\mathbf{M}$  of  $\mathbf{M}^*$  using the data samples in hand, and compute the corresponding eigenvectors or singular vectors of this surrogate matrix.

Viewed in this light, this monograph aims to identify key factors—e.g., certain spectral structure of  $\mathbf{M}^*$  as well as the size of the approximation error  $\mathbf{M} - \mathbf{M}^*$ —that exert main influences on the efficacy of the resultant spectral methods.

## 1.2 A modern statistical perspective

The idea of spectral methods can be traced back to early statistical literature on methods of moments (e.g., Pearson (1894) and Hansen (1982)), where one seeks to extract key parameters of the probability distributions of interest by examining the empirical moments of data. While classical matrix perturbation theory lays a sensible foundations for the analysis of spectral methods (Stewart and Sun, 1990), the theoretical understanding can be considerably enhanced through the lens of statistical modeling—a way of thinking that has flourished in the past decade. To the best of our knowledge, however, a systematic and comprehensive introduction to the modern statistical foundation of spectral methods, as well as an overview of recent advances, is previously unavailable.

The current monograph aims to fill this gap by developing a coherent and accessible treatment of spectral methods from a modern statistical perspective. Highlighting algorithmic implications that inform practice, our exposition gravitates around the following central questions: how to characterize the sample efficiency of spectral methods in reaching a prescribed accuracy level, and how to assess the stability of spectral methods in the face of random noise, missing data, and adversarial corruptions? We underscore several distinguishing features of our treatment compared to prior studies:

- In comparison to the worst-case performance guarantees derived solely based on classical matrix perturbation theory, our statistical treatment emphasizes the benefit of harnessing the “typical” behavior of data models, which offers key insights into how to harvest performance gains by leveraging intrinsic properties of data generating mechanisms.
- In contrast to classical asymptotic theory (Van der Vaart, 2000), we adopt a non-asymptotic (or finite-sample) analysis framework that draws on tools from recent developments of concentration inequalities (Tropp, 2015) and high-dimensional statistics (Wainwright, 2019). This framework accommodates the scenario where both the sample size and the number of features are enormous, and

unveils a clearer and more complete picture about the interplay and trade-off between salient model parameters.

Another unique feature of this monograph is a principled introduction of *fine-grained entrywise analysis* (e.g., a theory studying  $\ell_\infty$  eigenvector perturbation), which reflects cutting-edge research activities in this area. This is particularly important when, for example, demonstrating the feasibility of exact clustering or perfect ranking in the aforementioned applications. In truth, an effective entrywise analysis framework cannot be readily obtained from classical matrix analysis alone, and has only recently become available owing to the emergence of modern statistical toolboxes. In particular, we shall present a powerful framework, called *leave-one-out analysis*, that proves effective and versatile for delivering fine-grained performance guarantees for spectral methods in a variety of problems.

### 1.3 Organization

We now present a high-level overview of the structure of this monograph.

- Chapter 2 reviews the fundamentals of classical matrix perturbation theory for spectral analysis, focusing on  $\ell_2$ -type distances measured by the spectral norm and the Frobenius norm. This chapter covers the celebrated Davis-Kahan  $\sin \Theta$  theorem for eigenspace perturbation, the Wedin theorem for singular subspace perturbation, and an extension to probability transition matrices, laying the algebraic foundations for the remaining chapters.
- Chapter 3 explores the utility of  $\ell_2$  matrix perturbation theory when paired with statistical tools, presenting a unified recipe for statistical analysis empowered by non-asymptotic matrix tail bounds. We develop spectral methods for a variety of statistical data science applications, and derive nearly tight theoretical guarantees (up to logarithmic factors) based on this unified recipe.
- Chapter 4 develops fine-grained perturbation theory for spectral analysis in terms of  $\ell_\infty$  and  $\ell_{2,\infty}$  metrics, based on a leave-one-out analysis framework rooted in probability theory. Its effectiveness

is demonstrated through concrete applications including community recovery and matrix completion. This analysis framework also enables a non-asymptotic distributional theory for spectral methods, which paves the way for uncertainty quantification in applications like noisy matrix completion.

- Chapter 5 concludes this monograph by identifying a few directions that are worthy of future investigation.

While this monograph pursues a coherent and accessible treatment that might appeal to a broad audience, it does not necessarily deliver the sharpest possible results for the applications discussed herein in terms of the logarithmic terms and/or pre-constants. The bibliographic notes at the end of each chapter contain information about the state-of-the-art theory for each application as a pointer to further readings.

#### **1.4 What is not here and complementary readings**

The topics presented in this monograph do not cover the tensor decomposition methods studied in another recent strand of work (Anandkumar *et al.*, 2014). While such tensor-based methods are also sometimes referred to as spectral methods, their primary focus is to invoke tensor decomposition to learn latent variables, based on higher-order moments estimated from data samples. We elect not to discuss this class of methods but instead refer the interested reader to the recently published monograph by Janzamin *et al.* (2019). Another monograph by Kannan and Vempala (2009) provides an in-depth computational and algorithmic treatment of spectral methods from the perspective of theoretical computer science. The applications and results covered therein (e.g., fast matrix multiplication) complement the ones presented in the current monograph. In addition, spectral methods have been frequently employed to initialize nonconvex optimization algorithms. We will not elaborate on the nonconvex optimization aspect here but instead recommend the reader to the recent overview article by Chi *et al.* (2019). Finally, spectral methods are widely adopted to estimate high-dimensional covariance and precision matrices, and extract latent factors for econometric and

statistical modeling. This topic alone has a huge literature, and we refer the interested reader to Fan *et al.* (2020b) for in-depth discussions.

## 1.5 Notation

Before moving forward, let us introduce some notation that will be used throughout this monograph.

First of all, we reserve boldfaced symbols for vectors, matrices and tensors. For any matrix  $\mathbf{A}$ , let  $\sigma_j(\mathbf{A})$  (resp.  $\lambda_j(\mathbf{A})$ ) represent its  $j$ -th largest singular value (resp. eigenvalue). In particular,  $\sigma_{\max}(\mathbf{A})$  (resp.  $\lambda_{\max}(\mathbf{A})$ ) stands for the largest singular value (resp. eigenvalue) of  $\mathbf{A}$ , while  $\sigma_{\min}(\mathbf{A})$  (resp.  $\lambda_{\min}(\mathbf{A})$ ) indicates the smallest singular value (resp. eigenvalue) of  $\mathbf{A}$ . We use  $\mathbf{A}^\top$  to denote the transpose of  $\mathbf{A}$ , and let  $\mathbf{A}_{i\cdot}$  and  $\mathbf{A}_{\cdot i}$  indicate the  $i$ -th row and the  $i$ -th column of  $\mathbf{A}$ , respectively. We follow standard conventions by letting  $\mathbf{I}_n$  be the  $n \times n$  identity matrix,  $\mathbf{1}_n$  the  $n$ -dimensional all-one vector, and  $\mathbf{0}_n$  the  $n$ -dimensional all-zero vector; we shall often suppress the subscript as long as it is clear from the context. The  $i$ -th standard basis vector is denoted by  $\mathbf{e}_i$  throughout. The notation  $\mathcal{O}^{n \times r}$  ( $r \leq n$ ) represents the set of all  $n \times r$  orthonormal matrices (whose columns are orthonormal). Moreover, we refer to  $[n]$  as the set  $\{1, \dots, n\}$ .

Next, we turn to vector and matrix norms. For any vector  $\mathbf{v}$ , we denote by  $\|\mathbf{v}\|_2$ ,  $\|\mathbf{v}\|_1$  and  $\|\mathbf{v}\|_\infty$  its  $\ell_2$  norm,  $\ell_1$  norm and  $\ell_\infty$  norm, respectively. For any matrix  $\mathbf{A} = [A_{i,j}]_{1 \leq i \leq m, 1 \leq j \leq n}$ , we let  $\|\mathbf{A}\|$ ,  $\|\mathbf{A}\|_*$ ,  $\|\mathbf{A}\|_F$  and  $\|\mathbf{A}\|_\infty$  represent respectively its spectral norm (i.e., the largest singular value of  $\mathbf{A}$ ), its nuclear norm (i.e., the sum of singular values of  $\mathbf{A}$ ), its Frobenius norm (i.e.,  $\|\mathbf{A}\|_F := \sqrt{\sum_{i,j} A_{i,j}^2}$ ), and its entrywise  $\ell_\infty$  norm (i.e.,  $\|\mathbf{A}\|_\infty := \max_{i,j} |A_{i,j}|$ ). We also refer to  $\|\mathbf{A}\|_{2,\infty}$  as the  $\ell_{2,\infty}$  norm of  $\mathbf{A}$ , defined as  $\|\mathbf{A}\|_{2,\infty} := \max_i \|\mathbf{A}_{i\cdot}\|_2$ . Similarly, we define the  $\ell_{\infty,2}$  norm of  $\mathbf{A}$  as  $\|\mathbf{A}\|_{\infty,2} := \|\mathbf{A}^\top\|_{2,\infty}$ . In addition, for any matrices  $\mathbf{A} = [A_{i,j}]_{1 \leq i \leq m, 1 \leq j \leq n}$  and  $\mathbf{B} = [B_{i,j}]_{1 \leq i \leq m, 1 \leq j \leq n}$ , the inner product of  $\mathbf{A}$  and  $\mathbf{B}$  is defined as and denoted by  $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{1 \leq i \leq m, 1 \leq j \leq n} A_{i,j} B_{i,j} = \text{Tr}(\mathbf{A}^\top \mathbf{B})$ .

When it comes to diagonal matrices, we employ  $\text{diag}([\theta_1, \theta_2, \dots, \theta_r])$  to abbreviate the diagonal matrix with diagonal elements  $\theta_1, \dots, \theta_r$ . For any diagonal matrix  $\Theta = \text{diag}([\theta_1, \theta_2, \dots, \theta_r])$ , we adopt the shorthand

notation  $\sin \Theta := \text{diag}([\sin \theta_1, \sin \theta_2, \dots, \sin \theta_r])$ ; the notation  $\sin^2 \Theta$ ,  $\cos \Theta$ , and  $\cos^2 \Theta$  is defined analogously.

Finally, this monograph makes heavy use of the following standard notation: (1)  $f(n) = O(g(n))$  or  $f(n) \lesssim g(n)$  means that there exists a universal constant  $c > 0$  such that  $|f(n)| \leq c|g(n)|$  holds for all sufficiently large  $n$ ; (2)  $f(n) \gtrsim g(n)$  means that there exists a universal constant  $c > 0$  such that  $|f(n)| \geq c|g(n)|$  holds for all sufficiently large  $n$ ; (3)  $f(n) \asymp g(n)$  means that there exist universal constants  $c_1, c_2 > 0$  such that  $c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|$  holds for all sufficiently large  $n$ ; and (4)  $f(n) = o(g(n))$  indicates that  $f(n)/g(n) \rightarrow 0$  as  $n \rightarrow \infty$ . Additionally, we sometimes use  $f(n) \gg g(n)$  (resp.  $f(n) \ll g(n)$ ) to indicate that there exists some sufficiently large (resp. small) universal constant  $c > 0$  such that  $|f(n)| \geq c|g(n)|$  (resp.  $|f(n)| \leq c|g(n)|$ ).

## References

---

- Abbe, E. (2017). “Community detection and stochastic block models: recent developments”. *Journal of Machine Learning Research*. 18(1): 6446–6531.
- Abbe, E., A. Bandeira, and G. Hall. (2016). “Exact recovery in the stochastic block model”. *IEEE Transactions on Information Theory*. 62(1): 471–487.
- Abbe, E., J. Fan, and K. Wang. (2020a). “An  $L_p$  theory of PCA and spectral clustering”. *arXiv preprint arXiv:2006.14062*.
- Abbe, E., J. Fan, K. Wang, and Y. Zhong. (2020b). “Entrywise eigenvector analysis of random matrices with low expected rank”. *The Annals of Statistics*. 48(3): 1452–1474.
- Achlioptas, D. and F. McSherry. (2007). “Fast computation of low-rank matrix approximations”. *Journal of the ACM*. 54(2): 9–es.
- Agarwal, A., S. Kakade, and L. F. Yang. (2020). “Model-based reinforcement learning with a generative model is minimax optimal”. In: *Conference on Learning Theory*. 67–83.
- Agarwal, A., P. Patil, and S. Agarwal. (2018). “Accelerated spectral ranking”. In: *International Conference on Machine Learning*. 70–79.
- Agterberg, J., Z. Lubberts, and C. Priebe. (2021). “Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence”. *arXiv preprint arXiv:2105.13346*.



- Ahmed, A., B. Recht, and J. Romberg. (2013). “Blind deconvolution using convex programming”. *IEEE Transactions on Information Theory*. 60(3): 1711–1732.
- Ahn, K., K. Lee, and C. Suh. (2018). “Hypergraph spectral clustering in the weighted stochastic block model”. *IEEE Journal of Selected Topics in Signal Processing*. 12(5): 959–974.
- Ahn, S. C. and A. R. Horenstein. (2013). “Eigenvalue ratio test for the number of factors”. *Econometrica*. 81(3): 1203–1227.
- Aloise, D., A. Deshpande, P. Hansen, and P. Popat. (2009). “NP-hardness of Euclidean sum-of-squares clustering”. *Machine learning*. 75(2): 245–248.
- Alon, N., M. Krivelevich, and B. Sudakov. (1998). “Finding a large hidden clique in a random graph”. *Random Structures & Algorithms*. 13(3-4): 457–466.
- Amini, A. A. and M. J. Wainwright. (2008). “High-dimensional analysis of semidefinite relaxations for sparse principal components”. In: *International Symposium on Information Theory*. 2454–2458.
- Amini, A. A., A. Chen, P. J. Bickel, and E. Levina. (2013). “Pseudo-likelihood methods for community detection in large sparse networks”. *The Annals of Statistics*. 41(4): 2097–2122.
- Amjad, M., D. Shah, and D. Shen. (2018). “Robust synthetic control”. *The Journal of Machine Learning Research*. 19(1): 802–852.
- Anandkumar, A., R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. (2014). “Tensor decompositions for learning latent variable models”. *Journal of Machine Learning Research*. 15: 2773–2832.
- Anderson, T. W. (1962). *An introduction to multivariate statistical analysis*. Wiley New York.
- Awasthi, P., A. S. Bandeira, M. Charikar, R. Krishnaswamy, S. Villar, and R. Ward. (2015). “Relax, no need to round: Integrality of clustering formulations”. In: *Conference on Innovations in Theoretical Computer Science*. 191–200.
- Awasthi, P. and O. Sheffet. (2012). “Improved spectral-norm bounds for clustering”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer. 37–49.

- Azaouzi, M., D. Rhouma, and L. B. Romdhane. (2019). “Community detection in large-scale social networks: state-of-the-art and future directions”. *Social Network Analysis and Mining*. 9(1): 23.
- Bahmani, S. and J. Romberg. (2017). “Phase retrieval meets statistical learning theory: A flexible convex relaxation”. In: *Artificial Intelligence and Statistics*. 252–260.
- Bai, J. (2003). “Inferential theory for factor models of large dimensions”. *Econometrica*. 71(1): 135–171.
- Bai, J. (2009). “Panel data models with interactive fixed effects”. *Econometrica*. 77(4): 1229–1279.
- Bai, J. and S. Ng. (2002). “Determining the number of factors in approximate factor models”. *Econometrica*. 70(1): 191–221.
- Bai, J. and S. Ng. (2008). “Forecasting economic time series using targeted predictors”. *Journal of Econometrics*. 146(2): 304–317.
- Bai, J. and P. Wang. (2016). “Econometric analysis of large factor models”. *Annual Review of Economics*. 8: 53–80.
- Bair, E., T. Hastie, D. Paul, and R. Tibshirani. (2006). “Prediction by supervised principal components”. *Journal of the American Statistical Association*. 101(473): 119–137.
- Bajaj, C., T. Gao, Z. He, Q. Huang, and Z. Liang. (2018). “SMAC: Simultaneous mapping and clustering using spectral decompositions”. In: *International Conference on Machine Learning*. 324–333.
- Balakrishnan, S., M. J. Wainwright, and B. Yu. (2017). “Statistical guarantees for the EM algorithm: From population to sample-based analysis”. *The Annals of Statistics*. 45(1): 77–120.
- Balakrishnan, S., M. Xu, A. Krishnamurthy, and A. Singh. (2011). “Noise thresholds for spectral clustering”. In: *Advances in Neural Information Processing Systems*. 954–962.
- Balzano, L., Y. Chi, and Y. M. Lu. (2018). “Streaming PCA and subspace tracking: The missing data case”. *Proceedings of the IEEE*. 106(8): 1293–1310.
- Bandeira, A. S., N. Boumal, and A. Singer. (2017). “Tightness of the maximum likelihood semidefinite relaxation for angular synchronization”. *Mathematical Programming*. 163(1-2): 145–167.

- Bandeira, A. S. and R. Van Handel. (2016). “Sharp nonasymptotic bounds on the norm of random matrices with independent entries”. *The Annals of Probability*. 44(4): 2479–2506.
- Bao, Z., X. Ding, J. Wang, and K. Wang. (2020). “Statistical inference for principal components of spiked covariance matrix”. *arXiv preprint arXiv:2008.11903*.
- Bao, Z., X. Ding, and K. Wang. (2021). “Singular vector and singular subspace distribution for the matrix denoising model”. *The Annals of Statistics*. 49(1): 370–392.
- Barak, B. and A. Moitra. (2016). “Noisy tensor completion via the sum-of-squares hierarchy”. In: *Conference on Learning Theory*. PMLR. 417–445.
- Belkin, M. and P. Niyogi. (2003). “Laplacian eigenmaps for dimensionality reduction and data representation”. *Neural Computation*. 15(6): 1373–1396.
- Bhatia, R. (2013). *Matrix analysis*. Vol. 169. Springer Science & Business Media.
- Binkiewicz, N., J. T. Vogelstein, and K. Rohe. (2017). “Covariate-assisted spectral clustering”. *Biometrika*. 104(2): 361–377.
- Boucheron, S., G. Lugosi, and P. Massart. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.
- Boumal, N. and P.-A. Absil. (2015). “Low-rank matrix completion via preconditioned optimization on the Grassmann manifold”. *Linear Algebra and its Applications*. 475: 200–239.
- Bradley, R. A. and M. E. Terry. (1952). “Rank analysis of incomplete block designs: I. The method of paired comparisons”. *Biometrika*. 39(3/4): 324–345.
- Brémaud, P. (2013). *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Vol. 31. Springer Science & Business Media.
- Browet, A., P.-A. Absil, and P. Van Dooren. (2011). “Community detection for hierarchical image segmentation”. In: *International Workshop on Combinatorial Image Analysis*. 358–371.
- Cai, C., G. Li, Y. Chi, H. V. Poor, and Y. Chen. (2021a). “Subspace estimation from unbalanced and incomplete data matrices:  $\ell_{2,\infty}$  statistical guarantees”. *The Annals of Statistics*. 49(2): 944–967.

- Cai, C., G. Li, H. V. Poor, and Y. Chen. (2019a). “Nonconvex low-rank symmetric tensor completion from noisy data”. In: *Advances in Neural Information Processing Systems*. 1863–1874.
- Cai, C., G. Li, H. V. Poor, and Y. Chen. (2021b). “Nonconvex low-rank tensor completion from noisy data”. *Operations Research*.
- Cai, C., H. V. Poor, and Y. Chen. (2020). “Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality”. In: *International Conference on Machine Learning*. 1271–1282.
- Cai, J.-F., H. Liu, and Y. Wang. (2019b). “Fast rank-one alternating minimization algorithm for phase retrieval”. *Journal of Scientific Computing*. 79(1): 128–147.
- Cai, T. T. and X. Li. (2015). “Robust and computationally feasible community detection in the presence of arbitrary outlier nodes”. *The Annals of Statistics*. 43(3): 1027–1059.
- Cai, T. T., X. Li, and Z. Ma. (2016). “Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow”. *The Annals of Statistics*. 44(5): 2221–2251.
- Cai, T. T., Z. Ma, and Y. Wu. (2013). “Sparse PCA: Optimal rates and adaptive estimation”. *The Annals of Statistics*. 41(6): 3074–3110.
- Cai, T. T. and A. Zhang. (2015). “ROP: Matrix recovery via rank-one projections”. *The Annals of Statistics*. 43(1): 102–138.
- Cai, T. T. and A. Zhang. (2018). “Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics”. *The Annals of Statistics*. 46(1): 60–89.
- Cambareri, V. and L. Jacques. (2016). “A non-convex blind calibration method for randomised sensing strategies”. In: *2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*. 16–20.
- Candès, E., X. Li, and M. Soltanolkotabi. (2015a). “Phase retrieval via Wirtinger flow: Theory and algorithms”. *IEEE Transactions on Information Theory*. 61(4): 1985–2007.
- Candès, E. J. (2014). “Mathematics of sparsity (and a few other things)”. In: *Proceedings of the International Congress of Mathematicians, Seoul, South Korea*. Vol. 123.

- Candès, E. J., X. Li, Y. Ma, and J. Wright. (2011). “Robust principal component analysis?” *Journal of the ACM (JACM)*. 58(3): 1–37.
- Candès, E. J., X. Li, and M. Soltanolkotabi. (2015b). “Phase retrieval from coded diffraction patterns”. *Applied and Computational Harmonic Analysis*. 39(2): 277–299.
- Candès, E. J. and Y. Plan. (2010). “Matrix completion with noise”. *Proceedings of the IEEE*. 98(6): 925–936.
- Candès, E. J., T. Strohmer, and V. Voroninski. (2013). “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming”. *Communications on Pure and Applied Mathematics*. 66(8): 1241–1274.
- Candès, E. J. and B. Recht. (2009). “Exact matrix completion via convex optimization”. *Foundations of Computational Mathematics*. 9(6): 717–772.
- Candès, E. J. and T. Tao. (2010). “The power of convex relaxation: Near-optimal matrix completion”. *IEEE Transactions on Information Theory*. 56(5): 2053–2080.
- Cao, Y. and Y. Xie. (2015). “Poisson matrix recovery and completion”. *IEEE Transactions on Signal Processing*. 64(6): 1609–1620.
- Cape, J., M. Tang, and C. E. Priebe. (2019). “The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics”. *The Annals of Statistics*. 47(5): 2405–2439.
- Caplin, A. and B. Nalebuff. (1991). “Aggregation and social choice: A mean voter theorem”. *Econometrica: Journal of the Econometric Society*: 1–23.
- Chandra, R., T. Goldstein, and C. Studer. (2019). “Phasepack: A phase retrieval library”. In: *2019 13th International conference on Sampling Theory and Applications (SampTA)*. IEEE. 1–5.
- Chandrasekaran, V., P. A. Parrilo, and A. S. Willsky. (2010). “Latent variable graphical model selection via convex optimization”. In: *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 1610–1613.
- Chandrasekaran, V., S. Sanghavi, P. A. Parrilo, and A. S. Willsky. (2011). “Rank-sparsity incoherence for matrix decomposition”. *SIAM Journal on Optimization*. 21(2): 572–596.

- Charisopoulos, V., Y. Chen, D. Davis, M. Diaz, L. Ding, and D. Drusvyatskiy. (2021). “Low-rank matrix recovery with composite optimization: good conditioning and rapid convergence”. *Foundations of Computational Mathematics*: 1–89.
- Charisopoulos, V., D. Davis, M. Diaz, and D. Drusvyatskiy. (2019). “Composite optimization for robust blind deconvolution”. *arXiv preprint arXiv:1901.01624*.
- Chatterjee, S. (2014). “Matrix estimation by universal singular value thresholding”. *The Annals of Statistics*. 43(1): 177–214.
- Chaudhuri, K., F. Chung, and A. Tsiatas. (2012). “Spectral clustering of graphs with general degrees in the extended planted partition model”. In: *Conference on Learning Theory*. 35.1–35.23.
- Chen, P.-H. C., J. Chen, Y. Yeshurun, U. Hasson, J. Haxby, and P. J. Ramadge. (2015a). “A reduced-dimension fMRI shared response model”. In: *Advances in Neural Information Processing Systems*. 460–468.
- Chen, J., D. Liu, and X. Li. (2020a). “Nonconvex rectangular matrix completion via gradient descent without  $\ell_{2,\infty}$  regularization”. *IEEE Transactions on Information Theory*. 66(9): 5806–5841.
- Chen, L. H., L. Goldstein, and Q.-M. Shao. (2010). *Normal approximation by Stein’s method*. Springer Science & Business Media.
- Chen, P., A. Fannjiang, and G.-R. Liu. (2017). “Phase retrieval by linear algebra”. *SIAM Journal on Matrix Analysis and Applications*. 38(3): 854–868.
- Chen, P.-Y. and A. O. Hero. (2015). “Phase transitions in spectral community detection”. *IEEE Transactions on Signal Processing*. 63(16): 4339–4347.
- Chen, P., C. Gao, and A. Y. Zhang. (2020b). “Partial recovery for top- $K$  ranking: Optimality of MLE and sub-optimality of spectral method”. *arXiv preprint arXiv:2006.16485*.
- Chen, S., S. Liu, and Z. Ma. (2020c). “Global and individualized community detection in inhomogeneous multilayer networks”. *arXiv preprint arXiv:2012.00933*.
- Chen, X., P. N. Bennett, K. Collins-Thompson, and E. Horvitz. (2013). “Pairwise ranking aggregation in a crowdsourced setting”. In: *International conference on Web search and data mining*. 193–202.

- Chen, X. and Y. Yang. (2020). “Cutoff for exact recovery of Gaussian mixture models”. *arXiv preprint arXiv:2001.01194*.
- Chen, Y., Y. Chi, and A. Goldsmith. (2015b). “Exact and stable covariance estimation from quadratic sampling via convex programming”. *IEEE Transactions on Information Theory*. 61(7): 4034–4059.
- Chen, Y., C. Ma, H. V. Poor, and Y. Chen. (2021a). “Learning mixtures of low-rank models”. *IEEE Transactions on Information Theory*. 67(7): 4613–4636.
- Chen, Y. and Y. Chi. (2018). “Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization”. *IEEE Signal Processing Magazine*. 35(4): 14–31.
- Chen, Y. and M. J. Wainwright. (2015). “Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees”. *arXiv preprint arXiv:1509.03025*.
- Chen, Y. and E. Candès. (2017). “Solving random quadratic systems of equations is nearly as easy as solving linear systems”. *Communications on Pure and Applied Mathematics*. 70(5): 822–883.
- Chen, Y. and E. Candès. (2018). “The projected power method: An efficient algorithm for joint alignment from pairwise differences”. *Communications on Pure and Applied Mathematics*. 71(8): 1648–1714.
- Chen, Y., C. Cheng, and J. Fan. (2021b). “Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”. *The Annals of Statistics*. 49(1): 435–458.
- Chen, Y. and Y. Chi. (2014). “Robust spectral compressed sensing via structured matrix completion”. *IEEE Transactions on Information Theory*. 60(10): 6576–6601.
- Chen, Y., Y. Chi, J. Fan, and C. Ma. (2019a). “Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval”. *Mathematical Programming*. 176(1-2): 5–37.
- Chen, Y., Y. Chi, J. Fan, C. Ma, and Y. Yan. (2020d). “Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization”. *SIAM Journal on Optimization*. 30(4): 3098–3121.

- Chen, Y., J. Fan, C. Ma, and K. Wang. (2019b). “Spectral method and regularized MLE are both optimal for top- $K$  ranking”. *The Annals of Statistics*. 47(4): 2204–2235.
- Chen, Y., J. Fan, C. Ma, and Y. Yan. (2019c). “Inference and uncertainty quantification for noisy matrix completion”. *Proceedings of the National Academy of Sciences*. 116(46): 22931–22937.
- Chen, Y., J. Fan, C. Ma, and Y. Yan. (2021c). “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”. *accepted to The Annals of Statistics*.
- Chen, Y., J. Fan, B. Wang, and Y. Yan. (2021d). “Convex and nonconvex optimization are both minimax-optimal for noisy blind deconvolution”. *accepted to Journal of the American Statistical Association*.
- Chen, Y., L. J. Guibas, and Q.-X. Huang. (2014). “Near-optimal joint object matching via convex relaxation”. In: *International Conference on Machine Learning (ICML)*. 100–108.
- Chen, Y., G. Kamath, C. Suh, and D. Tse. (2016a). “Community recovery in graphs with locality”. In: *International Conference on Machine Learning*. 689–698.
- Chen, Y. and C. Suh. (2015). “Spectral MLE: Top- $K$  rank aggregation from pairwise comparisons”. In: *International Conference on Machine Learning*. 371–380.
- Chen, Y., C. Suh, and A. J. Goldsmith. (2016b). “Information recovery from pairwise measurements”. *IEEE Transactions on Information Theory*. 62(10): 5881–5905.
- Cheng, C., Y. Wei, and Y. Chen. (2021). “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”. *accepted to IEEE Transactions on Information Theory*.
- Cherapanamjeri, Y., K. Gupta, and P. Jain. (2017). “Nearly optimal robust matrix completion”. In: *International Conference on Machine Learning*. 797–805.
- Chernozhukov, V., C. Hansen, Y. Liao, and Y. Zhu. (2021). “Inference for low-rank models”. *arXiv preprint arXiv:2107.02602*.
- Chi, Y. and H. Fu. (2017). “Subspace learning from bits”. *IEEE Transactions on Signal Processing*. 65(17): 4429–4442.



- Chi, Y., Y. M. Lu, and Y. Chen. (2019). “Nonconvex optimization meets low-rank matrix factorization: An overview”. *IEEE Transactions on Signal Processing*. 67(20): 5239–5269.
- Chin, P., A. Rao, and V. Vu. (2015). “Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery”. In: *Conference on Learning Theory*. 391–423.
- Coja-Oghlan, A. (2010). “Graph partitioning via adaptive spectral techniques”. *Combinatorics, Probability & Computing*. 19(2): 227.
- Cole, S. and Y. Zhu. (2020). “Exact recovery in the hypergraph stochastic block model: A spectral algorithm”. *Linear Algebra and its Applications*. 593: 45–73.
- Cook, R. D. (2007). “Fisher lecture: Dimension reduction in regression”. *Statistical Science*. 22(1): 1–26.
- Cui, Q. and L. F. Yang. (2021). “Minimax Sample Complexity for Turn-based Stochastic Game”. In: *Conference on Uncertainty in Artificial Intelligence*.
- Dalvi, N., A. Dasgupta, R. Kumar, and V. Rastogi. (2013). “Aggregating crowdsourced binary ratings”. In: *International conference on World Wide Web*. 285–294.
- Dan, C., Y. Wei, and P. Ravikumar. (2020). “Sharp Statistical Guarantees for Adversarially Robust Gaussian Classification”. In: *International Conference on Machine Learning*. 2345–2355.
- Dasgupta, S. (1999). “Learning mixtures of Gaussians”. In: *40th Annual Symposium on Foundations of Computer Science*. IEEE. 634–644.
- Davenport, M. A., Y. Plan, E. Van Den Berg, and M. Wootters. (2014). “1-bit matrix completion”. *Information and Inference: A Journal of the IMA*. 3(3): 189–223.
- Davenport, M. A. and J. Romberg. (2016). “An overview of low-rank matrix recovery from incomplete observations”. *IEEE Journal of Selected Topics in Signal Processing*. 10(4): 608–622.
- Davis, C. and W. M. Kahan. (1970). “The rotation of eigenvectors by a perturbation. III”. *SIAM Journal on Numerical Analysis*. 7(1): 1–46.

- Dhifallah, O., C. Thrampoulidis, and Y. M. Lu. (2017). “Phase retrieval via linear programming: Fundamental limits and algorithmic improvements”. In: *2017 55th Annual Allerton Conference on Communication, Control, and Computing*. IEEE. 1071–1077.
- Diaconis, P. and L. Saloff-Coste. (1993). “Comparison theorems for reversible Markov chains”. *The Annals of Applied Probability*: 696–730.
- Ding, L. and Y. Chen. (2020). “Leave-one-out approach for matrix completion: Primal and dual analysis”. *IEEE Transactions on Information Theory*. 66(11): 7274–7301.
- Ding, X. (2020). “High dimensional deformed rectangular matrices with applications in matrix denoising”. *Bernoulli*. 26(1): 387–417.
- Dong, J. and Y. Shi. (2018). “Nonconvex demixing from bilinear measurements”. *IEEE Transactions on Signal Processing*. 66(19): 5152–5166.
- Dopico, F. M. (2000). “A note on  $\sin \Theta$  theorems for singular subspace variations”. *BIT Numerical Mathematics*. 40(2): 395–403.
- Du, S. S., W. Hu, S. M. Kakade, J. D. Lee, and Q. Lei. (2020). “Few-shot learning via learning the representation, provably”. In: *International Conference on Learning Representations*.
- Duan, N. and K.-C. Li. (1991). “Slicing regression: a link-free regression method”. *The Annals of Statistics*: 505–530.
- Duan, Y., T. Ke, and M. Wang. (2019). “State aggregation learning from Markov transition data”. In: *Advances in Neural Information Processing Systems*. 4486–4495.
- Duchi, J. C. and F. Ruan. (2019). “Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval”. *Information and Inference: A Journal of the IMA*. 8(3): 471–529.
- Dudeja, R., M. Bakhshizadeh, J. Ma, and A. Maleki. (2020). “Analysis of spectral methods for phase retrieval with random orthogonal matrices”. *IEEE Transactions on Information Theory*. 66(8): 5182–5203.
- Durrett, R. (2007). *Random graph dynamics*. Vol. 200. No. 7.
- Dwork, C., R. Kumar, M. Naor, and D. Sivakumar. (2001). “Rank aggregation methods for the web”. In: *International conference on World Wide Web*. 613–622.

- El Karoui, N. (2018). “On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators”. *Probability Theory and Related Fields*. 170(1-2): 95–175.
- El Karoui, N., D. Bean, P. J. Bickel, C. Lim, and B. Yu. (2013). “On robust regression with high-dimensional predictors”. *Proceedings of the National Academy of Sciences*. 110(36): 14557–14562.
- Eldar, Y. C. and S. Mendelson. (2014). “Phase retrieval: Stability and recovery guarantees”. *Applied and Computational Harmonic Analysis*. 36(3): 473–494.
- Eldridge, J., M. Belkin, and Y. Wang. (2018). “Unperturbed: Spectral analysis beyond Davis-Kahan”. In: *Algorithmic Learning Theory*. 321–358.
- Eriksson, B., L. Balzano, and R. Nowak. (2012). “High-rank matrix completion”. In: *Artificial Intelligence and Statistics*. 373–381.
- Fan, J. and Y. Fan. (2008). “High dimensional classification using features annealed independence rules”. *Annals of statistics*. 36(6): 2605.
- Fan, J., Y. Fan, X. Han, and J. Lv. (2019a). “SIMPLE: Statistical inference on membership profiles in large networks”. *arXiv preprint arXiv:1910.01734*.
- Fan, J., Y. Fan, X. Han, and J. Lv. (2020a). “Asymptotic theory of eigenvectors for random matrices with diverging spikes”. *Journal of the American Statistical Association*: 1–63.
- Fan, J., F. Han, and H. Liu. (2014). “Challenges of big data analysis”. *National science review*. 1(2): 293–314.
- Fan, J., K. Li, and Y. Liao. (2021a). “Recent developments on factor models and applications in econometric learning”. *Annual Review of Financial Economics*. 13: 1–35.
- Fan, J., R. Li, C.-H. Zhang, and H. Zou. (2020b). *Statistical foundations of data science*. CRC press.
- Fan, J., Y. Liao, and M. Mincheva. (2013). “Large covariance estimation by thresholding principal orthogonal complements”. *Journal of the Royal Statistical Society. Series B, Statistical methodology*. 75(4): 603–680.

- Fan, J., Y. Liao, and W. Wang. (2016). “Projected principal component analysis in factor models”. *The Annals of Statistics*. 44(1): 219–254.
- Fan, J., Y. Liao, and J. Yao. (2015). “Power enhancement in high-dimensional cross-sectional tests”. *Econometrica*. 83(4): 1497–1541.
- Fan, J. and J. Lv. (2008). “Sure independence screening for ultrahigh dimensional feature space”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 70(5): 849–911.
- Fan, J., Q. Sun, W.-X. Zhou, and Z. Zhu. (2018a). “Principal component analysis for big data”. In: *Wiley StatsRef: Statistics Reference Online*. Wiley Online Library. 1–13.
- Fan, J., D. Wang, K. Wang, and Z. Zhu. (2019b). “Distributed estimation of principal eigenspaces”. *The Annals of Statistics*. 47(6): 3009–3031.
- Fan, J., K. Wang, Y. Zhong, and Z. Zhu. (2021b). “Robust high-dimensional factor models with applications to statistical machine learning”. *Statistical Science*. 36(2): 303–327.
- Fan, J., W. Wang, and Y. Zhong. (2018b). “An  $\ell_\infty$  eigenvector perturbation bound and its application to robust covariance estimation”. *Journal of Machine Learning Research*. 18: 1–42.
- Fan, Z., I. M. Johnstone, and Y. Sun. (2018c). “Spiked covariances and principal components analysis in high-dimensional random effects models”. *arXiv preprint arXiv:1806.09529*.
- Fannjiang, A. and T. Strohmer. (2020). “The numerics of phase retrieval”. *Acta Numerica*. 29: 125–228.
- Fei, Y. and Y. Chen. (2018a). “Exponential error rates of SDP for block models: Beyond Grothendieck’s inequality”. *IEEE Transactions on Information Theory*. 65(1): 551–571.
- Fei, Y. and Y. Chen. (2018b). “Hidden integrality of SDP relaxations for sub-Gaussian mixture models”. In: *Conference On Learning Theory*. PMLR. 1931–1965.
- Fei, Y. and Y. Chen. (2019). “Achieving the Bayes error rate in stochastic block model by SDP, robustly”. In: *Conference on Learning Theory*. 1235–1269.
- Ferguson, T. S. (1996). *A course in large sample theory*. Vol. 38. Chapman & Hall/CRC.
- Fienup, J. R. (1982). “Phase retrieval algorithms: A comparison.” *Applied optics*. 21(15): 2758–2769.

- Fishkind, D. E., D. L. Sussman, M. Tang, J. T. Vogelstein, and C. E. Priebe. (2013). “Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown”. *SIAM Journal on Matrix Analysis and Applications*. 34(1): 23–39.
- Florescu, L. and W. Perkins. (2016). “Spectral thresholds in the bipartite stochastic block model”. In: *Conference on Learning Theory*. 943–959.
- Ford Jr, L. R. (1957). “Solution of a ranking problem from binary comparisons”. *The American Mathematical Monthly*. 64(8P2): 28–33.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin. (2000). “The generalized dynamic-factor model: Identification and estimation”. *Review of Economics and statistics*. 82(4): 540–554.
- Fortunato, S. and D. Hric. (2016). “Community detection in networks: A user guide”. *Physics Reports*. 659: 1–44.
- Foucart, S., D. Needell, R. Pathak, Y. Plan, and M. Wootters. (2020). “Weighted matrix completion from non-random, non-uniform sampling patterns”. *IEEE Transactions on Information Theory*. 67(2): 1264–1290.
- Fu, H., Y. Chi, and Y. Liang. (2020). “Guaranteed recovery of one-hidden-layer neural networks via cross entropy”. *IEEE Transactions on Signal Processing*. 68: 3225–3235.
- Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou. (2017). “Achieving optimal misclassification proportion in stochastic block models”. *Journal of Machine Learning Research*. 18(1): 1980–2024.
- Ge, R., C. Jin, P. Netrapalli, and A. Sidford. (2016). “Efficient algorithms for large-scale generalized eigenvector computation and canonical correlation analysis”. In: *International Conference on Machine Learning*. 2741–2750.
- Georghiadis, A. S., P. N. Belhumeur, and D. J. Kriegman. (2001). “From few to many: Illumination cone models for face recognition under variable lighting and pose”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23(6): 643–660.
- Ghosh, A., S. Kale, and P. McAfee. (2011). “Who moderates the moderators? crowdsourcing abuse detection in user-generated content”. In: *ACM conference on Electronic commerce*. 167–176.

- Ghosh, A. and R. Kannan. (2020). “Alternating minimization converges super-linearly for mixed linear regression”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 1093–1103.
- Giraud, C. and N. Verzelen. (2019). “Partial recovery bounds for clustering with the relaxed  $K$ -means”. *Mathematical Statistics and Learning*. 1(3): 317–374.
- Goldstein, T. and C. Studer. (2018). “Phasemax: Convex phase retrieval via basis pursuit”. *IEEE Transactions on Information Theory*. 64(4): 2675–2689.
- Gross, D. (2011). “Recovering low-rank matrices from few coefficients in any basis”. *IEEE Transactions on Information Theory*. 57(3): 1548–1566.
- Gross, D., F. Kraemer, and R. Kueng. (2017). “Improved recovery guarantees for phase retrieval from coded diffraction patterns”. *Applied and Computational Harmonic Analysis*. 42(1): 37–64.
- Hajek, B., S. Oh, and J. Xu. (2014). “Minimax-optimal inference from partial rankings”. *Advances in Neural Information Processing Systems*. 27: 1475–1483.
- Hajek, B., Y. Wu, and J. Xu. (2016). “Achieving exact cluster recovery threshold via semidefinite programming”. *IEEE Transactions on Information Theory*. 62(5): 2788–2797.
- Han, R., R. Willett, and A. Zhang. (2020). “An optimal statistical and computational framework for generalized tensor estimation”. *arXiv preprint arXiv:2002.11255*.
- Han, X., X. Tong, and Y. Fan. (2021). “Eigen selection in spectral clustering: a theory guided practice”. *Journal of the American Statistical Association*. (just-accepted): 1–33.
- Han, X., Q. Yang, and Y. Fan. (2019). “Universal rank inference via residual subsampling with application to large networks”. *arXiv preprint arXiv:1912.11583*.
- Hand, P. (2017). “Phaselift is robust to a constant fraction of arbitrary errors”. *Applied and Computational Harmonic Analysis*. 42(3): 550–562.
- Hand, P. and V. Voroninski. (2016). “Corruption robust phase retrieval via linear programming”. *arXiv preprint arXiv:1612.03547*.

- Hansen, L. P. (1982). “Large sample properties of generalized method of moments estimators”. *Econometrica: Journal of the Econometric Society*: 1029–1054.
- Hardt, M. (2014). “Understanding alternating minimization for matrix completion”. In: *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*. IEEE. 651–660.
- Hillar, C. J. and L.-H. Lim. (2013). “Most tensor problems are NP-hard”. *Journal of the ACM*. 60(6): 1–39.
- Holland, P. W., K. B. Laskey, and S. Leinhardt. (1983). “Stochastic blockmodels: First steps”. *Social networks*. 5(2): 109–137.
- Hopkins, S. B., T. Schramm, J. Shi, and D. Steurer. (2016). “Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors”. In: *Symposium on Theory of Computing*. 178–191.
- Horn, R. A. and C. R. Johnson. (2012). *Matrix analysis*. Cambridge university press.
- Howard, S. R., A. Ramdas, J. McAuliffe, and J. Sekhon. (2020). “Time-uniform Chernoff bounds via nonnegative supermartingales”. *Probability Surveys*. 17: 257–317.
- Hsu, D. (2016). “COMS 4772: advanced machine learning”. *Lecture notes, Columbia University*.
- Hsu, D. and S. M. Kakade. (2013). “Learning mixtures of spherical Gaussians: Moment methods and spectral decompositions”. In: *Conference on Innovations in Theoretical Computer Science*. 11–20.
- Huang, Q., Z. Liang, H. Wang, S. Zuo, and C. Bajaj. (2019a). “Tensor maps for synchronizing heterogeneous shape collections”. *ACM Transactions on Graphics (TOG)*. 38(4): 1–18.
- Huang, W. and P. Hand. (2018). “Blind deconvolution by a steepest descent algorithm on a quotient manifold”. *SIAM Journal on Imaging Sciences*. 11(4): 2757–2785.
- Huang, X., Z. Liang, X. Zhou, Y. Xie, L. J. Guibas, and Q. Huang. (2019b). “Learning transformation synchronization”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 8082–8091.
- Huang, Q.-X. and L. Guibas. (2013). “Consistent shape maps via semidefinite programming”. In: *Computer Graphics Forum*. Vol. 32. No. 5. 177–186.

- Huber, P. J. (2004). *Robust statistics*. Vol. 523. John Wiley & Sons.
- Hunter, D. R. (2004). “MM algorithms for generalized Bradley-Terry models”. *The annals of statistics*. 32(1): 384–406.
- Iguchi, T., D. G. Mixon, J. Peterson, and S. Villar. (2017). “Probably certifiably correct  $k$ -means clustering”. *Mathematical Programming*. 165(2): 605–642.
- Jaganathan, K., Y. C. Eldar, and B. Hassibi. (2016). “Phase retrieval: An overview of recent developments”. *Optical Compressive Imaging*: 263–296.
- Jagatap, G. and C. Hegde. (2019). “Sample-efficient algorithms for recovering structured signals from magnitude-only measurements”. *IEEE Transactions on Information Theory*. 65(7): 4434–4456.
- Jain, P. and P. Kar. (2017). “Non-convex optimization for machine learning”. *Foundations and Trends® in Machine Learning*. 10(3-4): 142–363.
- Jain, P., P. Netrapalli, and S. Sanghavi. (2013). “Low-rank matrix completion using alternating minimization”. In: *Symposium on Theory of Computing*. 665–674.
- Jalali, A., Y. Chen, S. Sanghavi, and H. Xu. (2011). “Clustering partially observed graphs via convex optimization”. In: *International Conference on Machine Learning*. 1001–1008.
- Jang, M., S. Kim, C. Suh, and S. Oh. (2016). “Top- $K$  ranking from pairwise comparisons: When spectral ranking is optimal”. *arXiv preprint arXiv:1603.04153*.
- Janzamin, M., R. Ge, J. Kossaifi, and A. Anandkumar. (2019). “Spectral learning on matrices and tensors”. *Foundations and Trends in Machine Learning*. 12(5-6): 393–536.
- Javanmard, A. and A. Montanari. (2013). “Localization from incomplete noisy distance measurements”. *Foundations of Computational Mathematics*. 13(3): 297–345.
- Javanmard, A. and A. Montanari. (2018). “Debiasing the Lasso: Optimal sample size for Gaussian designs”. *The Annals of Statistics*. 46(6A): 2593–2622.
- Jeong, H. and C. S. Güntürk. (2017). “Convergence of the randomized Kaczmarz method for phase retrieval”. *arXiv preprint arXiv:1706.10291*.



- Jiang, S., Y. Liu, and Y. Tang. (2017). “A unified perturbation analysis framework for countable Markov chains”. *Linear Algebra and its Applications*. 529: 413–440.
- Jin, C., S. M. Kakade, and P. Netrapalli. (2016a). “Provable efficient on-line matrix completion via non-convex stochastic gradient descent”. In: *Advances in Neural Information Processing Systems*. 4520–4528.
- Jin, C., Y. Zhang, S. Balakrishnan, M. J. Wainwright, and M. I. Jordan. (2016b). “Local maxima in the likelihood of Gaussian mixture models: Structural results and algorithmic consequences”. In: *Advances in Neural Information Processing Systems*. 4116–4124.
- Jin, J. (2015). “Fast community detection by SCORE”. *The Annals of Statistics*. 43(1): 57–89.
- Jin, J., Z. T. Ke, and W. Wang. (2017). “Phase transitions for high dimensional clustering and related problems”. *The Annals of Statistics*. 45(5): 2151–2189.
- Johnstone, I. M. (2001). “On the distribution of the largest eigenvalue in principal components analysis”. *The Annals of statistics*. 29(2): 295–327.
- Johnstone, I. M. and A. Y. Lu. (2009). “On consistency and sparsity for principal components analysis in high dimensions”. *Journal of the American Statistical Association*. 104(486): 682–693.
- Johnstone, I. M. and D. Paul. (2018). “PCA in high dimensions: An orientation”. *Proceedings of the IEEE*. 106(8): 1277–1292.
- Jolliffe, I. T. (1986). “Principal components in regression analysis”. In: *Principal component analysis*. Springer. 129–155.
- Kalai, A. T., A. Moitra, and G. Valiant. (2010). “Efficiently learning mixtures of two Gaussians”. In: *Proceedings of the forty-second ACM symposium on Theory of computing*. 553–562.
- Kannan, R., H. Salmasian, and S. Vempala. (2008). “The spectral method for general mixture models”. *SIAM Journal on Computing*. 38(3): 1141–1156.
- Kannan, R. and S. Vempala. (2009). *Spectral algorithms*. Now Publishers Inc.

- Karger, D. R., S. Oh, and D. Shah. (2013). “Efficient crowdsourcing for multi-class labeling”. In: *ACM SIGMETRICS/international conference on Measurement and modeling of computer systems*. 81–92.
- Karger, D. R., S. Oh, and D. Shah. (2014). “Budget-optimal task allocation for reliable crowdsourcing systems”. *Operations Research*. 62(1): 1–24.
- Kato, T. (2013). *Perturbation theory for linear operators*. Vol. 132. Springer Science & Business Media.
- Ke, Z. T. and M. Wang. (2017). “A new SVD approach to optimal topic estimation”. *arXiv preprint arXiv:1704.07016*.
- Keshavan, R., A. Montanari, and S. Oh. (2009). “Matrix completion from noisy entries”. In: *Advances in Neural Information Processing Systems*. 952–960.
- Keshavan, R. H., A. Montanari, and S. Oh. (2010). “Matrix completion from a few entries”. *IEEE Transactions on Information Theory*. 56(6): 2980–2998.
- Klopp, O. (2014). “Noisy low-rank matrix completion with general sampling distribution”. *Bernoulli*. 20(1): 282–303.
- Kolda, T. G. and B. W. Bader. (2009). “Tensor decompositions and applications”. *SIAM review*. 51(3): 455–500.
- Koltchinskii, V., M. Loffler, and R. Nickl. (2020). “Efficient estimation of linear functionals of principal components”. *The Annals of Statistics*. 48(1): 464–490.
- Koltchinskii, V. and K. Lounici. (2016). “Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 52. No. 4. 1976–2013.
- Koltchinskii, V., K. Lounici, and A. B. Tsybakov. (2011). “Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion”. *The Annals of Statistics*. 39(5): 2302–2329.
- Koltchinskii, V. and D. Xia. (2016). “Perturbation of linear forms of singular vectors under Gaussian noise”. In: *High Dimensional Probability VII*. Springer. 397–423.

- Kong, W., R. Somani, Z. Song, S. Kakade, and S. Oh. (2020). “Meta-learning for mixed linear regression”. In: *International Conference on Machine Learning*. 5394–5404.
- Krahmer, F. and D. Stöger. (2021). “On the convex geometry of blind deconvolution and matrix completion”. *Communications on Pure and Applied Mathematics*. 74(4): 790–832.
- Kreimer, N., A. Stanton, and M. D. Sacchi. (2013). “Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction”. *Geophysics*. 78(6): V273–V284.
- Kumar, A. and R. Kannan. (2010). “Clustering with spectral norm and the  $k$ -means algorithm”. In: *Proceedings of 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 299–308.
- Kwon, J., N. Ho, and C. Caramanis. (2021). “On the minimax optimality of the EM algorithm for learning two-component mixed linear regression”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 1405–1413.
- Lawley, D. N. and A. E. Maxwell. (1962). “Factor analysis as a statistical method”. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 12(3): 209–229.
- Le, C. M. and E. Levina. (2015). “Estimating the number of communities in networks by spectral methods”. *arXiv preprint arXiv:1507.00827*.
- Le, C. M., E. Levina, and R. Vershynin. (2017). “Concentration and regularization of random graphs”. *Random Structures & Algorithms*. 51(3): 538–561.
- Le, C. M., E. Levina, and R. Vershynin. (2018). “Concentration of random graphs and application to community detection”. In: *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific. 2925–2943.
- Lee, K., F. Krahmer, and J. Romberg. (2018). “Spectral methods for passive imaging: Nonasymptotic performance and robustness”. *SIAM Journal on Imaging Sciences*. 11(3): 2110–2164.
- Lee, K., Y. Wu, and Y. Bresler. (2017). “Near-optimal compressed sensing of a class of sparse low-rank matrices via sparse power factorization”. *IEEE Transactions on Information Theory*. 64(3): 1666–1698.

- Lei, J. and A. Rinaldo. (2015). “Consistency of spectral clustering in stochastic block models”. *The Annals of Statistics*. 43(1): 215–237.
- Lei, J. and L. Zhu. (2014). “A generic sample splitting approach for refined community recovery in stochastic block models”. *arXiv preprint arXiv:1411.1469*.
- Lei, L. (2019). “Unified  $\ell_{2 \rightarrow \infty}$  eigenspace perturbation theory for symmetric random matrices”. *arXiv preprint arXiv:1909.04798*.
- Lei, L., P. J. Bickel, and N. El Karoui. (2018). “Asymptotics for high dimensional regression M-estimates: Fixed design results”. *Probability Theory and Related Fields*. 172(3-4): 983–1079.
- Li, B. (2018). *Sufficient dimension reduction: Methods and applications with R*. CRC Press.
- Li, G., C. Cai, Y. Gu, H. V. Poor, and Y. Chen. (2021a). “Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps”. *arXiv preprint arXiv:2104.03298*.
- Li, G. and Y. Gu. (2021). “Theory of spectral method for union of subspaces-based random geometry graph”. In: *International Conference on Machine Learning*. PMLR. 6337–6345.
- Li, G., Y. Wei, Y. Chi, Y. Gu, and Y. Chen. (2020a). “Breaking the sample size barrier in model-based reinforcement learning with a generative model”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 12861–12872.
- Li, J. and A. O. Hero. (2004). “A fast spectral method for active 3D shape reconstruction”. *Journal of Mathematical Imaging and Vision*. 20(1-2): 73–87.
- Li, K.-C. (1992). “On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma”. *Journal of the American Statistical Association*. 87(420): 1025–1039.
- Li, R.-C. (1998). “Relative perturbation theory: II. Eigenspace and singular subspace variations”. *SIAM Journal on Matrix Analysis and Applications*. 20(2): 471–492.
- Li, X. and V. Voroninski. (2013). “Sparse signal recovery from quadratic measurements via convex programming”. *SIAM Journal on Mathematical Analysis*. 45(5): 3019–3033.
- Li, X., Y. Chen, and J. Xu. (2021b). “Convex relaxation methods for community detection”. *Statistical Science*. 36(1): 2–15.

- Li, X., Y. Li, S. Ling, T. Strohmer, and K. Wei. (2020b). “When do birds of a feather flock together?  $k$ -means, proximity, and conic programming”. *Mathematical Programming*. 179(1-2): 295–341.
- Li, X., S. Ling, T. Strohmer, and K. Wei. (2019). “Rapid, robust, and reliable blind deconvolution via nonconvex optimization”. *Applied and Computational Harmonic Analysis*. 47(3): 893–934.
- Li, Y., K. Lee, and Y. Bresler. (2018). “Blind gain and phase calibration via sparse spectral methods”. *IEEE Transactions on Information Theory*. 65(5): 3097–3123.
- Li, Y., Y. Chi, H. Zhang, and Y. Liang. (2020c). “Non-convex low-rank matrix recovery with arbitrary outliers via median-truncated gradient descent”. *Information and Inference: A Journal of the IMA*. 9(2): 289–325.
- Li, Y., C. Ma, Y. Chen, and Y. Chi. (2021c). “Nonconvex matrix factorization from rank-one measurements”. *IEEE Transactions on Information Theory*. 67(3): 1928–1950.
- Lin, N., R. Jing, Y. Wang, E. Yonekura, J. Fan, and L. Xue. (2017). “A statistical investigation of the dependence of tropical cyclone intensity change on the surrounding environment”. *Monthly Weather Review*. 145(7): 2813–2831.
- Ling, S. (2020). “Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods”. *arXiv preprint arXiv:2008.05341*.
- Ling, S. and T. Strohmer. (2019). “Regularized gradient descent: A non-convex recipe for fast joint blind deconvolution and demixing”. *Information and Inference: A Journal of the IMA*. 8(1): 1–49.
- Liu, A. and A. Moitra. (2020). “Tensor Completion Made Practical”. *Neural Information Processing Systems*.
- Liu, J., P. Musialski, P. Wonka, and J. Ye. (2012). “Tensor completion for estimating missing values in visual data”. *IEEE transactions on Pattern Analysis and Machine Intelligence*. 35(1): 208–220.
- Liu, L. T., E. Dobriban, and A. Singer. (2018). “ $e$  PCA: High dimensional exponential family PCA”. *The Annals of Applied Statistics*. 12(4): 2121–2150.

- Liu, Y. (2012). “Perturbation bounds for the stationary distributions of Markov chains”. *SIAM Journal on Matrix Analysis and Applications*. 33(4): 1057–1074.
- Lloyd, S. (1982). “Least squares quantization in PCM”. *IEEE Transactions on Information Theory*. 28(2): 129–137.
- Löffler, M., A. Y. Zhang, and H. H. Zhou. (2021). “Optimality of spectral clustering for Gaussian mixture model”. *accepted to The Annals of Statistics*.
- Loh, P.-L. and M. J. Wainwright. (2012). “High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity”. *The Annals of Statistics*. 40(3): 1637–1664.
- Lounici, K. (2014). “High-dimensional covariance matrix estimation with missing observations”. *Bernoulli*. 20(3): 1029–1058.
- Lu, Y. and H. H. Zhou. (2016). “Statistical and computational guarantees of Lloyd’s algorithm and its variants”. *arXiv preprint arXiv:1612.02099*.
- Lu, Y. M. and G. Li. (2020). “Phase transitions of spectral initialization for high-dimensional non-convex estimation”. *Information and Inference: A Journal of the IMA*. 9(3): 507–541.
- Luce, R. D. (2012). *Individual choice behavior: A theoretical analysis*. Courier Corporation.
- Luo, W., W. Alghamdi, and Y. M. Lu. (2019). “Optimal spectral initialization for signal recovery with applications to phase retrieval”. *IEEE Transactions on Signal Processing*. 67(9): 2347–2356.
- Lusseau, D., K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. (2003). “The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations”. *Behavioral Ecology and Sociobiology*. 54(4): 396–405.
- Ma, C., Y. Li, and Y. Chi. (2021a). “Beyond Procrustes: Balancing-free gradient descent for asymmetric low-rank matrix sensing”. *IEEE Transactions on Signal Processing*. 69: 867–877.
- Ma, C., K. Wang, Y. Chi, and Y. Chen. (2020). “Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution”. *Foundations of Computational Mathematics*. 20(3): 451–632.

- Ma, J., R. Dudgeja, J. Xu, A. Maleki, and X. Wang. (2021b). “Spectral method for phase retrieval: an expectation propagation perspective”. *IEEE Transactions on Information Theory*. 67(2): 1332–1355.
- Ma, J., J. Xu, and A. Maleki. (2019). “Optimization-based AMP for Phase Retrieval: The Impact of Initialization and  $\ell_2$ -regularization”. *IEEE Transactions on Information Theory*. 65(6): 3600–3629.
- Ma, Z. (2013). “Sparse principal component analysis and iterative thresholding”. *The Annals of Statistics*. 41(2): 772–801.
- MacQueen, J. (1967). “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. No. 14. Oakland, CA, USA. 281–297.
- Mahoney, M. W. (2016). “Lecture notes on randomized linear algebra”. *arXiv preprint arXiv:1608.04481*.
- Massoulié, L. (2014). “Community detection thresholds and the weak Ramanujan property”. In: *Symposium on Theory of computing*. 694–703.
- Maunu, T., T. Zhang, and G. Lerman. (2019). “A well-tempered landscape for non-convex robust subspace recovery”. *Journal of Machine Learning Research*. 20(37): 1–59.
- McCrae, R. R. and O. P. John. (1992). “An introduction to the five-factor model and its applications”. *Journal of personality*. 60(2): 175–215.
- McRae, A. D. and M. A. Davenport. (2021). “Low-rank matrix completion and denoising under Poisson noise”. *Information and Inference: A Journal of the IMA*. 10(2): 697–720.
- McSherry, F. (2001). “Spectral partitioning of random graphs”. In: *IEEE Symposium on Foundations of Computer Science*. 529–537.
- Mezard, M. and A. Montanari. (2009). *Information, Physics, and Computation*. Oxford University Press.
- Michoel, T. and B. Nachtergaele. (2012). “Alignment and integration of complex networks by hypergraph-based spectral clustering”. *Physical Review E*. 86(5): 056111.
- Mitrophanov, A. Y. (2005). “Sensitivity and convergence of uniformly ergodic Markov chains”. *Journal of Applied Probability*. 42(4): 1003–1014.

- Mixon, D. G., S. Villar, and R. Ward. (2017). “Clustering subgaussian mixtures by semidefinite programming”. *Information and Inference: A Journal of the IMA*. 6(4): 389–415.
- Moitra, A. and A. S. Wein. (2019). “Spectral methods from tensor networks”. In: *ACM SIGACT Symposium on Theory of Computing*. 926–937.
- Monardo, V. and Y. Chi. (2019). “On the sensitivity of spectral initialization for noisy phase retrieval”. In: *International Conference on Acoustics, Speech and Signal Processing*. 5172–5176.
- Mondelli, M. and A. Montanari. (2019). “Fundamental limits of weak recovery with applications to phase retrieval”. *Foundations of Computational Mathematics*. 19(3): 703–773.
- Montanari, A. (2011). “EE 378B: Statistical Signal Processing”. *Lecture notes, Stanford University*.
- Montanari, A. and N. Sun. (2018). “Spectral algorithms for tensor completion”. *Communications on Pure and Applied Mathematics*. 71(11): 2381–2425.
- Mossel, E., J. Neeman, and A. Sly. (2015). “Consistency thresholds for the planted bisection model”. In: *Symposium on Theory of computing*. 69–75.
- Nadler, B. (2008). “Finite sample approximation results for principal component analysis: A matrix perturbation approach”. *The Annals of Statistics*. 36(6): 2791–2817.
- Nayer, S., P. Narayanamurthy, and N. Vaswani. (2019). “Phaseless PCA: Low-rank matrix recovery from column-wise phaseless measurements”. In: *International Conference on Machine Learning*. 4762–4770.
- Ndaoud, M. (2018). “Sharp optimal recovery in the two component Gaussian mixture model”. *arXiv preprint arXiv:1812.08078*.
- Negahban, S. and M. J. Wainwright. (2011). “Estimation of (near) low-rank matrices with noise and high-dimensional scaling”. *The Annals of Statistics*. 39(2): 1069–1097.
- Negahban, S., S. Oh, and D. Shah. (2016). “Rank centrality: Ranking from pairwise comparisons”. *Operations Research*. 65(1): 266–287.



- Negahban, S. and M. J. Wainwright. (2012). “Restricted strong convexity and weighted matrix completion: Optimal bounds with noise”. *Journal of Machine Learning Research*. 13(1): 1665–1697.
- Netrapalli, P., P. Jain, and S. Sanghavi. (2015). “Phase retrieval using alternating minimization”. *IEEE Transactions on Signal Processing*. 18(63): 4814–4826.
- Netrapalli, P., U. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain. (2014). “Non-convex robust PCA”. In: *Advances in Neural Information Processing Systems*. 1107–1115.
- Newman, M. E. (2006). “Finding community structure in networks using the eigenvectors of matrices”. *Physical review E*. 74(3): 036104.
- Newman, M. E. (2013). “Spectral methods for community detection and graph partitioning”. *Physical Review E*. 88(4): 042822.
- Ng, A. Y., M. I. Jordan, and Y. Weiss. (2002). “On spectral clustering: Analysis and an algorithm”. In: *Advances in Neural Information Processing Systems*. 849–856.
- Oh, S. and D. Shah. (2014). “Learning mixed multinomial logit model from ordinal data”. In: *Advances in Neural Information Processing Systems*. 595–603.
- Oliveira, R. I. (2009). “Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges”. *arXiv preprint arXiv:0911.0600*.
- Olver, P. J., C. Shakiban, and C. Shakiban. (2006). *Applied linear algebra*. Vol. 1. Springer.
- Oymak, S., A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. (2015). “Simultaneously structured models with application to sparse and low-rank matrices”. *IEEE Transactions on Information Theory*. 61(5): 2886–2908.
- Pachauri, D., R. Kondor, and V. Singh. (2013). “Solving the multi-way matching problem by permutation synchronization”. *Advances in Neural Information Processing Systems*. 26: 1860–1868.
- Page, L., S. Brin, R. Motwani, and T. Winograd. (1999). “The PageRank citation ranking: Bringing order to the web.” *Tech. rep.* Stanford InfoLab.

- Pananjady, A. and M. J. Wainwright. (2020). “Instance-dependent  $\ell_\infty$ -bounds for policy evaluation in tabular reinforcement learning”. *IEEE Transactions on Information Theory*. 67(1): 566–585.
- Paul, D. (2007). “Asymptotics of sample eigenstructure for a large dimensional spiked covariance model”. *Statistica Sinica*: 1617–1642.
- Pearson, K. (1894). “Contributions to the mathematical theory of evolution”. *Philosophical Transactions of the Royal Society of London. A*. 185: 71–110.
- Peng, J. and Y. Wei. (2007). “Approximating  $k$ -means-type clustering via semidefinite programming”. *SIAM Journal on Optimization*. 18(1): 186–205.
- Perry, A., A. S. Wein, A. S. Bandeira, and A. Moitra. (2018). “Optimality and sub-optimality of PCA I: Spiked random matrix models”. *The Annals of Statistics*. 46(5): 2416–2451.
- Qu, Q., Y. Zhang, Y. C. Eldar, and J. Wright. (2019). “Convolutional phase retrieval via gradient descent”. *IEEE Transactions on Information Theory*. 66(3): 1785–1821.
- Raginsky, M. and I. Sason. (2013). “Concentration of measure inequalities in information theory, communications, and coding”. *Foundations and Trends in Communications and Information Theory*. 10(1-2): 1–250.
- Rajkumar, A. and S. Agarwal. (2014). “A statistical convergence perspective of algorithms for rank aggregation from pairwise data”. In: *International Conference on Machine Learning*. 118–126.
- Reynolds, D. A. and R. C. Rose. (1995). “Robust text-independent speaker identification using Gaussian mixture speaker models”. *IEEE Transactions on Speech and Audio Processing*. 3(1): 72–83.
- Richard, E. and A. Montanari. (2014). “A statistical model for tensor PCA”. In: *Advances in Neural Information Processing Systems*. 2897–2905.
- Rohe, K., S. Chatterjee, and B. Yu. (2011). “Spectral clustering and the high-dimensional stochastic blockmodel”. *The Annals of Statistics*. 39(4): 1878–1915.
- Rudolf, D. and N. Schweizer. (2018). “Perturbation theory for Markov chains via Wasserstein distance”. *Bernoulli*. 24(4A): 2610–2639.

- Salehi, F., E. Abbasi, and B. Hassibi. (2018a). “A precise analysis of phasemax in phase retrieval”. In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 976–980.
- Salehi, F., E. Abbasi, and B. Hassibi. (2018b). “Learning without the phase: Regularized phasemax achieves optimal sample complexity”. In: *Advances in Neural Information Processing Systems*. 8641–8652.
- Sanghavi, S., R. Ward, and C. D. White. (2017). “The local convexity of solving systems of quadratic equations”. *Results in Mathematics*. 71(3-4): 569–608.
- Sarkar, P. and P. J. Bickel. (2015). “Role of normalization in spectral clustering for stochastic blockmodels”. *The Annals of Statistics*. 43(3): 962–990.
- Scharf, L. L. (1991). “The SVD and reduced rank signal processing”. *Signal Processing*. 25(2): 113–133.
- Schönemann, P. H. (1966). “A generalized solution of the orthogonal procrustes problem”. *Psychometrika*. 31(1): 1–10.
- Seeley, J. R. (1949). “The net of reciprocal influence. a problem in treating sociometric data”. *Canadian Journal of Experimental Psychology*. 3: 234.
- Shah, N., S. Balakrishnan, A. Guntuboyina, and M. Wainwright. (2016). “Stochastically transitive models for pairwise comparisons: Statistical and computational issues”. In: *International Conference on Machine Learning*. 11–20.
- Shah, N. B., S. Balakrishnan, and M. J. Wainwright. (2019). “Feeling the Bern: Adaptive estimators for Bernoulli probabilities of pairwise comparisons”. *IEEE Transactions on Information Theory*. 65(8): 4854–4874.
- Shah, N. B., J. K. Bradley, A. Parekh, M. Wainwright, and K. Ramchandran. (2013). “A case for ordinal peer-evaluation in MOOCs”. In: *NIPS Workshop on Data Driven Education*. 1–8.
- Shapiro, A., Y. Xie, and R. Zhang. (2018). “Matrix completion with deterministic pattern: A geometric perspective”. *IEEE Transactions on Signal Processing*. 67(4): 1088–1103.
- Shechtman, Y., A. Beck, and Y. C. Eldar. (2014). “GESPAR: Efficient phase retrieval of sparse signals”. *IEEE Transactions on Signal Processing*. 62(4): 928–938.

- Shechtman, Y., Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. (2015). “Phase retrieval with application to optical imaging: a contemporary overview”. *IEEE Signal Processing Magazine*. 32(3): 87–109.
- Shen, Y., Q. Huang, N. Srebro, and S. Sanghavi. (2016). “Normalized spectral map synchronization”. In: *Advances in Neural Information Processing Systems*. 4925–4933.
- Shi, J. and J. Malik. (2000). “Normalized cuts and image segmentation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(8): 888–905.
- Sidiropoulos, N. D., L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. (2017). “Tensor decomposition for signal processing and machine learning”. *IEEE Transactions on Signal Processing*. 65(13): 3551–3582.
- Singer, A. (2011). “Angular synchronization by eigenvectors and semidefinite programming”. *Applied and Computational Harmonic Analysis*. 30(1): 20–36.
- Singer, A. and Y. Shkolnisky. (2011). “Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming”. *SIAM Journal on Imaging Sciences*. 4(2): 543–572.
- Soltanolkotabi, M. (2019). “Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization”. *IEEE Transactions on Information Theory*. 65(4): 2374–2400.
- Srivastava, P. R., P. Sarkar, and G. A. Hanasusanto. (2019). “A robust spectral clustering algorithm for sub-Gaussian mixture models with outliers”. *arXiv preprint arXiv:1912.07546*.
- Stein, C. (1972). “A bound for the error in the normal approximation to the distribution of a sum of dependent random variables”. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*.
- Stewart, G. W. and J.-G. Sun. (1990). *Matrix perturbation theory*. Academic Press.

- Stock, J. H. and M. W. Watson. (2002). “Forecasting using principal components from a large number of predictors”. *Journal of the American statistical association*. 97(460): 1167–1179.
- Stock, J. H. and M. W. Watson. (2016). “Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics”. In: *Handbook of macroeconomics*. Vol. 2. Elsevier. 415–525.
- Sun, J. (1987). *Matrix perturbation analysis (in Chinese)*. Vol. 6. Science Press.
- Sun, R. and Z.-Q. Luo. (2016). “Guaranteed matrix completion via non-convex factorization”. *IEEE Transactions on Information Theory*. 62(11): 6535–6579.
- Sun, Y., Z. Liang, X. Huang, and Q. Huang. (2018). “Joint map and symmetry synchronization”. In: *European Conference on Computer Vision (ECCV)*. 251–264.
- Sun, Y., J. Zhuo, A. Mohan, and Q. Huang. (2019). “ $K$ -best transformation synchronization”. In: *IEEE International Conference on Computer Vision*. 10252–10261.
- Sur, P. and E. J. Candès. (2019). “A modern maximum-likelihood theory for high-dimensional logistic regression”. *Proceedings of the National Academy of Sciences*. 116(29): 14516–14525.
- Sur, P., Y. Chen, and E. J. Candès. (2019). “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”. *Probability Theory and Related Fields*. 175(1-2): 487–558.
- Tan, Y. S. and R. Vershynin. (2019). “Phase retrieval via randomized Kaczmarz: Theoretical guarantees”. *Information and Inference: A Journal of the IMA*. 8(1): 97–123.
- Tao, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics*. Providence, Rhode Island: American Mathematical Society.
- Titterton, D. M., A. F. Smith, and U. E. Makov. (1985). *Statistical analysis of finite mixture distributions*. Wiley.
- Tomasi, C. and T. Kanade. (1992). “Shape and motion from image streams under orthography: a factorization method”. *International Journal of Computer Vision*. 9(2): 137–154.

- Tong, T., C. Ma, and Y. Chi. (2021a). “Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent”. *Journal of Machine Learning Research*. 22(150): 1–63.
- Tong, T., C. Ma, and Y. Chi. (2021b). “Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number”. *IEEE Transactions on Signal Processing*. 69: 2396–2409.
- Tong, T., C. Ma, A. Prater-Bennette, E. Tripp, and Y. Chi. (2021c). “Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements”. *arXiv preprint arXiv:2104.14526*.
- Tripuraneni, N., C. Jin, and M. Jordan. (2021). “Provable meta-learning of linear representations”. In: *International Conference on Machine Learning*. PMLR. 10434–10443.
- Tropp, J. (2011). “Freedman’s inequality for matrix martingales”. *Electronic Communications in Probability*. 16: 262–270.
- Tropp, J. A. (2012). “User-friendly tail bounds for sums of random matrices”. *Foundations of Computational Mathematics*. 12(4): 389–434.
- Tropp, J. A. (2015). “An introduction to matrix concentration inequalities”. *Foundations and Trends® in Machine Learning*. 8(1-2): 1–230.
- Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer.
- Tu, S., R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht. (2016). “Low-rank solutions of linear matrix equations via procrustes flow”. In: *International Conference on Machine Learning*. 964–973.
- Turk, M. A. and A. P. Pentland. (1991). “Face recognition using eigenfaces”. *Conference on Computer Vision and Pattern Recognition*: 586–591.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*. Vol. 3. Cambridge university press.
- Vaswani, N. (2020). “Nonconvex structured phase retrieval: A focus on provably correct approaches”. *IEEE Signal Processing Magazine*. 37(5): 67–77.

- Vaswani, N., Y. Chi, and T. Bouwmans. (2018). “Rethinking PCA for modern data sets: Theory, algorithms, and applications”. *Proceedings of the IEEE*. 106(8): 1274–1276.
- Vaswani, N., S. Nayer, and Y. C. Eldar. (2017). “Low-rank phase retrieval”. *IEEE Transactions on Signal Processing*. 65(15): 4059–4074.
- Vempala, S. and G. Wang. (2004). “A spectral algorithm for learning mixture models”. *Journal of Computer and System Sciences*. 68(4): 841–860.
- Vershynin, R. (2017). *High-dimensional probability*. Cambridge University Press.
- Vidal, R., Y. Ma, and S. S. Sastry. (2016). *Generalized principal component analysis*. Springer.
- Vigna, S. (2016). “Spectral ranking”. *Network Science*. 4(4): 433–445.
- Von Luxburg, U. (2007). “A tutorial on spectral clustering”. *Statistics and Computing*. 17(4): 395–416.
- Vu, V. and J. Lei. (2012). “Minimax rates of estimation for sparse PCA in high dimensions”. In: *Artificial intelligence and statistics*. 1278–1286.
- Vu, V. Q. and J. Lei. (2013). “Minimax sparse principal subspace estimation in high dimensions”. *The Annals of Statistics*. 41(6): 2905–2947.
- Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. Vol. 48. Cambridge University Press.
- Wang, B., Y. Yan, and J. Fan. (2021a). “Sample-efficient reinforcement learning for linearly-parameterized MDPs with a generative model”. *arXiv preprint arXiv:2105.14016*.
- Wang, G., G. B. Giannakis, and Y. C. Eldar. (2018a). “Solving systems of random quadratic equations via truncated amplitude flow”. *IEEE Transactions on Information Theory*. 64(2): 773–794.
- Wang, G., L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen. (2018b). “Sparse phase retrieval via truncated amplitude flow”. *IEEE Transactions on Signal Processing*. 66(2): 479–491.
- Wang, H., J. Chen, and K. Wei. (2021b). “Entrywise convergence of Riemannian gradient method for low rank tensor completion via Tucker decomposition”. *arXiv preprint arXiv:2108.07899*.

- Wang, W. and J. Fan. (2017). “Asymptotics of empirical eigenstructure for high dimensional spiked covariance”. *The Annals of Statistics*. 45(3): 1342.
- Wang, Z., Y. Liang, and P. Ji. (2020). “Spectral algorithms for community detection in directed networks”. *Journal of Machine Learning Research*. 21(153): 1–45.
- Wedin, P.-Å. (1972). “Perturbation bounds in connection with singular value decomposition”. *BIT Numerical Mathematics*. 12(1): 99–111.
- Wei, K., J.-F. Cai, T. F. Chan, and S. Leung. (2016). “Guarantees of Riemannian optimization for low rank matrix recovery”. *SIAM Journal on Matrix Analysis and Applications*. 37(3): 1198–1222.
- Wright, J. and Y. Ma. (2020). *High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications*. Cambridge University Press.
- Wu, Y. and P. Yang. (2020). “Optimal estimation of Gaussian mixtures via denoised method of moments”. *The Annals of Statistics*. 48(4): 1981–2007.
- Xia, D. (2021). “Normal approximation and confidence region of singular subspaces”. *Electronic Journal of Statistics*. 15(2): 3798–3851.
- Xia, D. and M. Yuan. (2019). “On polynomial time methods for exact low-rank tensor completion”. *Foundations of Computational Mathematics*. 19(6): 1265–1313.
- Xia, D. and M. Yuan. (2021). “Statistical inferences of linear forms for noisy matrix completion”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 83(1): 58–77.
- Xia, D., M. Yuan, and C.-H. Zhang. (2021). “Statistically optimal and computationally efficient low rank tensor completion from noisy entries”. *The Annals of Statistics*. 49(1): 76–99.
- Xia, D., A. R. Zhang, and Y. Zhou. (2020). “Inference for low-rank tensors—No need to debias”. *arXiv preprint arXiv:2012.14844*.
- Xia, Y., H. Tong, W. K. Li, and L.-X. Zhu. (2002). “An adaptive estimation of dimension reduction space”. *Journal of Royal Statistical Society Series B*. 64: 363–410.
- Xie, L., Y. Xie, and G. V. Moustakides. (2020). “Sequential subspace change point detection”. *Sequential Analysis*. 39(3): 307–335.



- Xu, J., D. J. Hsu, and A. Maleki. (2016). “Global analysis of expectation maximization for mixtures of two Gaussians”. In: *Advances in Neural Information Processing Systems*. 2676–2684.
- Xu, J., A. Maleki, and K. R. Rad. (2019). “Consistent risk estimation in high-dimensional linear regression”. *arXiv preprint arXiv:1902.01753*.
- Xu, L. and M. I. Jordan. (1996). “On convergence properties of the EM algorithm for Gaussian mixtures”. *Neural Computation*. 8(1): 129–151.
- Yan, Y., Y. Chen, and J. Fan. (2021). “Inference for heteroskedastic PCA with missing data”. *arXiv preprint arXiv:2107.12365*.
- Yang, Z., Z. Wang, H. Liu, Y. Eldar, and T. Zhang. (2016). “Sparse nonlinear regression: Parameter estimation under nonconvexity”. In: *International Conference on Machine Learning*. 2472–2481.
- Yang, Z., L. F. Yang, E. X. Fang, T. Zhao, Z. Wang, and M. Neykov. (2019). “Misspecified nonconvex statistical optimization for sparse phase retrieval”. *Mathematical Programming*. 176(1-2): 545–571.
- Yi, X., C. Caramanis, and S. Sanghavi. (2014). “Alternating minimization for mixed linear regression”. In: *International Conference on Machine Learning*. 613–621.
- Yi, X., D. Park, Y. Chen, and C. Caramanis. (2016). “Fast algorithms for robust PCA via gradient descent”. In: *Advances in Neural Information Processing Systems*. 4152–4160.
- Yu, Y., T. Wang, and R. J. Samworth. (2015). “A useful variant of the Davis-Kahan theorem for statisticians”. *Biometrika*. 102(2): 315–323.
- Yuan, Y. and A. Qu. (2018). “Community detection with dependent connectivity”. *arXiv preprint arXiv:1812.06406*.
- Yuan, Z., H. Wang, and Q. Wang. (2019). “Phase retrieval via sparse Wirtinger flow”. *Journal of Computational and Applied Mathematics*. 355: 162–173.
- Zhang, A. Y. and H. H. Zhou. (2020). “Theoretical and computational guarantees of mean field variational inference for community detection”. *The Annals of Statistics*. 48(5): 2575–2598.
- Zhang, A., T. T. Cai, and Y. Wu. (2021). “Heteroskedastic PCA: Algorithm, optimality, and applications”. *accepted to The Annals of Statistics*.

- Zhang, A. and M. Wang. (2020). “Spectral state compression of Markov processes”. *IEEE Transactions on Information Theory*. 66(5): 3202–3231.
- Zhang, A. and D. Xia. (2018). “Tensor SVD: Statistical and computational limits”. *IEEE Transactions on Information Theory*. 64(11): 7311–7338.
- Zhang, A. R., Y. Luo, G. Raskutti, and M. Yuan. (2020a). “ISLET: Fast and optimal low-rank tensor regression via importance sketching”. *SIAM Journal on Mathematics of Data Science*. 2(2): 444–479.
- Zhang, H., Y. Chi, and Y. Liang. (2016). “Provable non-convex phase retrieval with outliers: Median truncated Wirtinger flow”. In: *International Conference on Machine Learning*. 1022–1031.
- Zhang, H., Y. Chi, and Y. Liang. (2018a). “Median-truncated nonconvex approach for phase retrieval with outliers”. *IEEE Transactions on Information Theory*. 64(11): 7287–7310.
- Zhang, H., Y. Zhou, Y. Liang, and Y. Chi. (2017). “A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms”. *The Journal of Machine Learning Research*. 18(1): 5164–5198.
- Zhang, K., S. Kakade, T. Basar, and L. Yang. (2020b). “Model-based multi-agent RL in zero-sum Markov games with near-optimal sample complexity”. *Advances in Neural Information Processing Systems*. 33.
- Zhang, L., G. Wang, G. B. Giannakis, and J. Chen. (2018b). “Compressive phase retrieval via reweighted amplitude flow”. *IEEE Transactions on Signal Processing*. 66(19): 5029–5040.
- Zhang, X., S. Du, and Q. Gu. (2018c). “Fast and sample efficient inductive matrix completion via multi-phase Procrustes flow”. In: *International Conference on Machine Learning*. 5756–5765.
- Zhang, X., L. Wang, Y. Yu, and Q. Gu. (2018d). “A primal-dual analysis of global optimality in nonconvex low-rank matrix recovery”. In: *International Conference on Machine Learning*. 5857–5866.
- Zhang, Y., E. Levina, and J. Zhu. (2020c). “Detecting overlapping communities in networks using spectral methods”. *SIAM Journal on Mathematics of Data Science*. 2(2): 265–283.

- Zhang, Y., X. Chen, D. Zhou, and M. I. Jordan. (2014). “Spectral methods meet EM: A provably optimal algorithm for crowdsourcing”. In: *Advances in Neural Information Processing Systems*. 1260–1268.
- Zhang, Y., Q. Qu, and J. Wright. (2020d). “From symmetry to geometry: Tractable nonconvex problems”. *arXiv preprint arXiv:2007.06753*.
- Zhao, Y., E. Levina, and J. Zhu. (2012). “Consistency of community detection in networks under degree-corrected stochastic block models”. *The Annals of Statistics*. 40(4): 2266–2292.
- Zheng, Q. and J. Lafferty. (2015). “A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements”. In: *Advances in Neural Information Processing Systems*. 109–117.
- Zheng, Q. and J. Lafferty. (2016). “Convergence analysis for rectangular matrix completion using Burer-Monteiro factorization and gradient descent”. *arXiv preprint arXiv:1605.07051*.
- Zhong, K., Z. Song, P. Jain, P. L. Bartlett, and I. S. Dhillon. (2017). “Recovery guarantees for one-hidden-layer neural networks”. In: *International Conference on Machine Learning*. 4140–4149.
- Zhong, Y. and N. Boumal. (2018). “Near-optimal bounds for phase synchronization”. *SIAM Journal on Optimization*. 28(2): 989–1016.
- Zhu, Z., Y. Wang, D. Robinson, D. Naiman, R. Vidal, and M. Tsakiris. (2018). “Dual principal component pursuit: Improved analysis and efficient algorithms”. In: *Advances in Neural Information Processing Systems*. 2171–2181.
- Zhu, Z., T. Wang, and R. J. Samworth. (2019). “High-dimensional principal component analysis with heterogeneous missingness”. *arXiv preprint arXiv:1906.12125*.
- Zou, H., T. Hastie, and R. Tibshirani. (2006). “Sparse principal component analysis”. *Journal of Computational and Graphical Statistics*. 15(2): 265–286.