Yuejie Chi

Low-Rank Matrix Completion

Imagine one observes a small subset of entries in a large matrix and aims to recover the entire matrix. Without a priori knowledge of the matrix, this problem is highly ill-posed. Fortunately, data matrices often exhibit low-dimensional structures that can be used effectively to regularize the solution space. The celebrated effectiveness of principal component analysis (PCA) in science and engineering suggests that most variability of real-world data can be accounted for by projecting the data onto a few directions known as the principal components. Correspondingly, the data matrix can be modeled as a low-rank matrix, at least approximately. Is it possible to complete a partially observed matrix if its rank, i.e., its maximum number of linearly independent row or column vectors, is small?

Low-rank matrix completion arises in a variety of applications in recommendation systems, computer vision, and signal processing. As a motivating example, consider users' ratings of products arranged in a rating matrix. Each rating may only be affected by a small number of factors—such as price, quality, and utility—and how they are reflected on the products' specifications and users' expectations. Naturally, this suggests that the rating matrix is low rank, since the numbers of users and products are much higher than the number of factors. Often, the rating matrix is sparsely observed, and it is of great interest to predict the missing ratings to make targeted recommendations.

Relevance

The theory and algorithms of low-rank matrix completion have been significantly expanded in the last decade with converging efforts from signal processing, applied mathematics, statistics, optimization, and machine learning. This lecture note provides an introductory exposition of some key results in this rapidly developing field.

Prerequisites

We expect the readers to be familiar with basic concepts in linear algebra, optimization, and probability.

Problem statement

Let $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ be a rank-r matrix, whose thin singular value decomposition (SVD) is given as

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}, \tag{1}$$

where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}, \boldsymbol{V} \in \mathbb{R}^{n_2 \times r}$ are composed of orthonormal columns, and $\boldsymbol{\Sigma}$ is an r-dimensional diagonal matrix with the singular values arranged in a nonincreasing order, i.e., $\sigma_1 \geq \cdots \geq \sigma_r > 0$. The "degrees of freedom" of $\boldsymbol{M}$ is $(n_1 + n_2 - r)r$, which is the total number of parameters we need to uniquely specify $\boldsymbol{M}$.

Assume we are given partial observations of $\boldsymbol{M}$ over an index set $\Omega \subset \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\}$. To concisely put it, define the observation operator $\mathcal{P}_{\Omega}: \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{n_1 \times n_2}$ as

$$[\mathcal{P}_{\Omega}(\boldsymbol{M})]_{ij} = \begin{cases} M_{ij}, & (i,j) \in \Omega \\ 0, & \text{otherwise} \end{cases}.$$

Our goal is to recover $\boldsymbol{M}$ from $\mathcal{P}_{\Omega}(\boldsymbol{M})$, when the number of observation $m = |\Omega| \ll n_1 n_2$ is much smaller than the number of entries in $\boldsymbol{M}$, under the assumption that $\boldsymbol{M}$ is low rank, i.e., $r \ll \min\{n_1, n_2\}$. For notational simplicity in the sequel, let $n = \max\{n_1, n_2\}$.

Solution

Which low-rank matrices can we complete?

To begin, we ask the following question: What kind of low-rank matrices can we complete? As motivation, consider the following 4×4 rank-1 matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$, given as

$$\boldsymbol{M}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{M}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

The matrix $\boldsymbol{M}_1$ is more difficult to complete, since most of its entries are zero, and we need to collect more measurements to make sure enough mass comes from its nonzero entries. In contrast, the mass of $\boldsymbol{M}_2$ is more uniformly distributed across all entries, making it easier to propagate information from one entry to another.

To put it differently, a low-rank matrix is easier to complete if its energy spreads evenly across different coordinates.

To put it differently, a low-rank matrix is easier to complete if its energy spreads evenly across different coordinates. This property is captured by the notion of *coherence* [1], which measures the alignment between the column/row spaces of the low-rank matrix with standard basis vectors. For a matrix $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ with orthonormal columns, let $\boldsymbol{P}_{U}$ be the orthogonal projection onto the column space of $\boldsymbol{U}$. The coherence parameter of $\boldsymbol{U}$ is defined as

Digital Object Identifier 10.1109/MSP.2018.2832197
Date of publication: 28 August 2018

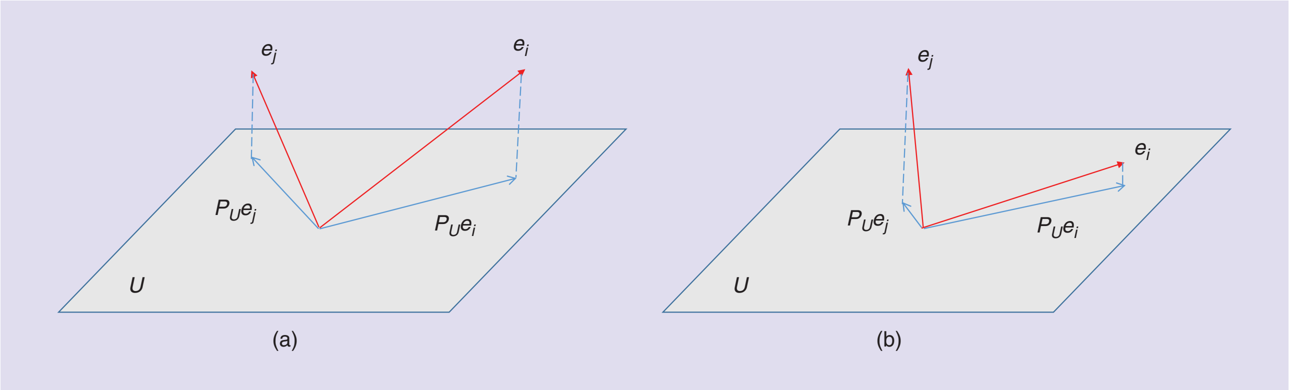

FIGURE 1. An illustration of the coherence parameter $\mu(\boldsymbol{U})$. $\mu(\boldsymbol{U})$ is small when all the standard basis vectors $\boldsymbol{e}_i$ have approximately the same projections onto the subspace $\boldsymbol{U}$, as shown in (a); $\mu(\boldsymbol{U})$ is large if $\boldsymbol{U}$ is too aligned with certain standard basis vector, as shown in (b).

$$\mu(\boldsymbol{U}) = \frac{n_1}{r}\max_{1\le i\le n_1}\|\boldsymbol{P}_U\boldsymbol{e}_i\|_2^2 = \frac{n_1}{r}\max_{1\le i\le n_1}\|\boldsymbol{U}^\top\boldsymbol{e}_i\|_2^2, \qquad (2)$$

where $\boldsymbol{e}_i$ is the ith standard basis vector. Figure 1 provides a geometric illustration of the coherence parameter $\mu(\boldsymbol{U})$.

For a low-matrix $\boldsymbol{M}$ whose SVD is given in (1), the coherence of $\boldsymbol{M}$ is defined as

$$\mu = \max\{\mu(\boldsymbol{U}),\mu(\boldsymbol{V})\}. \qquad (3)$$

Notably, the coherence μ is determined by the singular vectors of $\boldsymbol{M}$ and independent of its singular values. Since $1 \le \mu(\boldsymbol{U}) \le n_1/r$ and $1 \le \mu(\boldsymbol{V}) \le n_2/r$, we have $1 \le \mu \le n/r$. In the previous example, the coherence of $\boldsymbol{M}_1$ matches the upper bound n/r, while the coherence of $\boldsymbol{M}_2$ matches the lower bound one. The smaller μ is, the easier it is to complete the matrix.

Which observation patterns can we handle?

Low-rank matrix completion can still be hopeless even when most of the entries are revealed. Consider, for example, the following observation pattern for a 4×4 matrix:

$$\begin{bmatrix} \star & \star & \star & ? \\ \star & \star & \star & ? \\ \star & \star & \star & ? \\ \star & \star & \star & ? \end{bmatrix},$$

where $\star$ indicates an observed entry, and ? indicates a missing entry. The last column of the matrix cannot be recovered since it can lie anywhere in the column space of the low-rank matrix. Therefore, we require at least r observations per column/row. To bypass such pessimistic observation patterns, it is useful to think of random observation patterns. A popular choice is the Bernoulli model, where each entry is observed independently and identically with probability $p := m/(n_1n_2)$. By a coupon-collecting argument [2], under the Bernoulli model, it is impossible to recover a low-rank matrix with less than some constant times $\mu nr\log n$ measurements using any algorithm, which is referred to as the *information-theoretic lower bound.* Compared with the degrees of freedom, which is on the order of nr, we pay a price in sample complexity by a factor of $\mu\log n$, highlighting again the role of coherence in low-rank matrix completion.

Matrix completion via convex optimization

We present the first algorithm based on convex optimization. To promote the low-rank structure of the solution, a natural heuristic is to find the matrix with the minimum rank that is consistent with the observations, leading to

$$\min_{\boldsymbol{\Phi}\in\mathbb{R}^{n_1\times n_2}} \operatorname{rank}(\boldsymbol{\Phi}) \quad \text{s.t.} \quad \mathcal{P}_\Omega(\boldsymbol{\Phi}) = \mathcal{P}_\Omega(\boldsymbol{M}). \qquad (4)$$

However, since rank minimization is NP-hard, the above formulation is intractable. Motivated by the success of ℓ_1 norm minimization for sparse recovery in compressed sensing [3], we consider convex relaxation for the rank heuristic. Observing that the rank of $\boldsymbol{\Phi}$ equals to the number of its nonzero singular values, we replace rank $(\boldsymbol{\Phi})$ by the sum of its singular values, denoted as the nuclear norm:

$$\|\boldsymbol{\Phi}\|_* \triangleq \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\boldsymbol{\Phi}),$$

where $\sigma_i(\boldsymbol{\Phi})$ is the ith singular value of $\boldsymbol{\Phi}$. The nuclear norm is the tightest convex relaxation of the rank constraint, i.e., the nuclear norm ball $\{\boldsymbol{\Phi}:\|\boldsymbol{\Phi}\|_* \le 1\}$ is the convex hull of the collection of unit-norm rank-1 matrices: $\{\boldsymbol{u}\boldsymbol{v}^\top:\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1\}$. Notably, the nuclear norm is also unitarily invariant, and can be represented as the solution to a semidefinite program,

$$\|\boldsymbol{\Phi}\|_* = \min_{\boldsymbol{W}_1,\boldsymbol{W}_2} \frac{1}{2}(\operatorname{Tr}(\boldsymbol{W}_1)+\operatorname{Tr}(\boldsymbol{W}_2)) \quad \text{s.t.} \quad \begin{bmatrix}\boldsymbol{W}_1 & \boldsymbol{\Phi}\\ \boldsymbol{\Phi}^\top & \boldsymbol{W}_2\end{bmatrix} \succeq 0.$$

Hence, instead of solving (4) directly, we solve nuclear norm minimization, which searches for a matrix with the minimum nuclear norm that satisfies all the measurements:

$$\min_{\boldsymbol{\Phi}\in\mathbb{R}^{n_1\times n_2}} \|\boldsymbol{\Phi}\|_* \quad \text{s.t.} \quad \mathcal{P}_\Omega(\boldsymbol{\Phi}) = \mathcal{P}_\Omega(\boldsymbol{M}). \qquad (5)$$

This gives a convex program that can be solved efficiently in polynomial time. Moreover, it doesn't require knowledge of the rank a priori.

The performance of nuclear norm minimization has been investigated in a recent line of elegant works [2]–[5], which suggests it can exactly recover a low-rank matrix as soon as the number

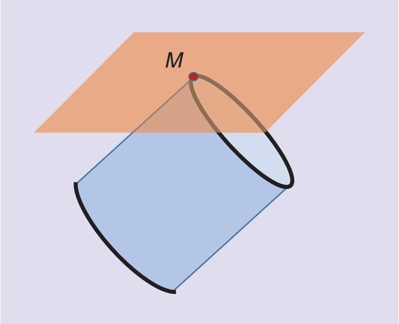

FIGURE 2. A geometric illustration of nuclear norm minimization: the cylinder represents level sets of the nuclear norm, and the hyperplane represents the measurement constraint. The two sets intersect at the thickened edges, which correspond to low-rank solutions.

of measurements is slightly larger than the information-theoretic lower bound by a logarithmic factor. Suppose that each entry of $\boldsymbol{M}$ is observed independently with probability $p \in (0,1)$. If p satisfies

$$p \geq C \frac{\mu r \log^2 n}{n},$$

for some large enough constant $C > 0$, then with high probability, the nuclear norm minimization algorithm (5) exactly recovers $\boldsymbol{M}$ as the unique optimal solution of (5). Figure 2 illustrates the geometry of nuclear norm minimization when the number of measurements is sufficiently large. When both μ and r are much smaller than n, this means we can recover a low-rank matrix even when the proportion of observations is vanishingly small.

Matrix completion via nonconvex optimization

The computational and memory complexities of nuclear norm minimization can be quite expensive for large-scale problems, even with first-order methods, due to optimizing over and storing the matrix variable $\boldsymbol{\Phi}$. Therefore, it is necessary to consider alternative approaches whose complexities scale more favorably in n. This leads to the second algorithm based on gradient descent using a proper initialization. If the rank of the matrix $\boldsymbol{M}$ is known, it is natural to incorporate this knowledge and consider a rank-constrained least-squares problem

$$\min_{\boldsymbol{\Phi} \in \mathbb{R}^{n_1 \times n_2}} \| \mathcal{P}_\Omega(\boldsymbol{\Phi} - \boldsymbol{M}) \|_{\mathrm{F}}^2, \quad \text{s.t rank } (\boldsymbol{\Phi}) \leq r, \tag{6}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix. Invoking the low-rank factorization $\boldsymbol{\Phi} = \boldsymbol{X}\boldsymbol{Y}^\top$, where $\boldsymbol{X} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{Y} \in \mathbb{R}^{n_2 \times r}$, we can rewrite (6) as an unconstrained, yet nonconvex optimization problem:

$$\min_{\boldsymbol{X},\boldsymbol{Y}} f(\boldsymbol{X},\boldsymbol{Y}) := \| \mathcal{P}_\Omega(\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{M}) \|_{\mathrm{F}}^2. \tag{7}$$

On one end, the memory complexities of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are linear in n instead of quadratic in n when dealing with $\boldsymbol{\Phi}$. On the other end, we can only determine $\boldsymbol{X}$ and $\boldsymbol{Y}$ up to invertible transforms in (7), since for any invertible matrix $\boldsymbol{Q} \in \mathbb{R}^{r \times r}$, we have $\boldsymbol{X}\boldsymbol{Y}^\top = (\boldsymbol{X}\boldsymbol{Q})(\boldsymbol{Y}\boldsymbol{Q}^{-\top})^\top$. To fix the scaling ambiguity, it is useful to consider a modified loss function

$$F(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{4p} f(\boldsymbol{X},\boldsymbol{Y}) + \frac{1}{16} \| X^\top \boldsymbol{X} - \boldsymbol{Y}^\top Y \|_{\mathrm{F}}^2,$$

where the second term is introduced to motivate solutions where $\boldsymbol{X}$ and $\boldsymbol{Y}$ have balanced norms. The observation probability p, if not known, can be faithfully estimated by the sample proportion $|\Omega|/(n_1 n_2)$.

How do we optimize the nonconvex loss $F(\boldsymbol{X},\boldsymbol{Y})$? A plausible strategy proceeds in two steps.

1) The first step aims to find an initialization that is close to the ground truth, which can be provided via the so-called spectral method [6]. Consider the partially observed matrix $(1/p)\mathcal{P}_\Omega(\boldsymbol{M})$, which is an unbiased estimate of $\boldsymbol{M}$ with expectation $\mathbb{E}[(1/p)\mathcal{P}_\Omega(\boldsymbol{M})] = \boldsymbol{M}$. Therefore, its best rank-r approximation produces a reasonably good initial guess. Let $\boldsymbol{U}_0 \boldsymbol{\Sigma}_0 \boldsymbol{V}_0^\top$ be the best rank-r approximation of $(1/p)\mathcal{P}_\Omega(\boldsymbol{M})$, where $\boldsymbol{U}_0 \in \mathbb{R}^{n_1 \times r}, \boldsymbol{V}_0 \in \mathbb{R}^{n_2 \times r}$ contain orthonormal columns and $\boldsymbol{\Sigma}_0$ is an $r \times r$ diagonal matrix. The spectral initialization sets $\boldsymbol{X}_0 = \boldsymbol{U}_0 \boldsymbol{\Sigma}_0^{1/2}$ and $\boldsymbol{Y}_0 = \boldsymbol{V}_0 \boldsymbol{\Sigma}_0^{1/2}$.
2) The second step aims to refine the initial estimate locally via simple iterative methods, such as gradient descent [7], [8], following the update rule

$$\begin{bmatrix} \boldsymbol{X}_{t+1} \\ \boldsymbol{Y}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \end{bmatrix} - \eta_t \begin{bmatrix} \nabla_{\boldsymbol{X}} F(\boldsymbol{X}_t, \boldsymbol{Y}_t) \\ \nabla_{\boldsymbol{Y}} F(\boldsymbol{X}_t, \boldsymbol{Y}_t) \end{bmatrix}, \tag{8}$$

where η_t is the step size, and $\nabla_{\boldsymbol{X}} F(\boldsymbol{X},\boldsymbol{Y}), \nabla_{\boldsymbol{Y}} F(\boldsymbol{X},\boldsymbol{Y})$ are the partial derivatives with respect to $\boldsymbol{X}$ and $\boldsymbol{Y}$ that can be derived easily.

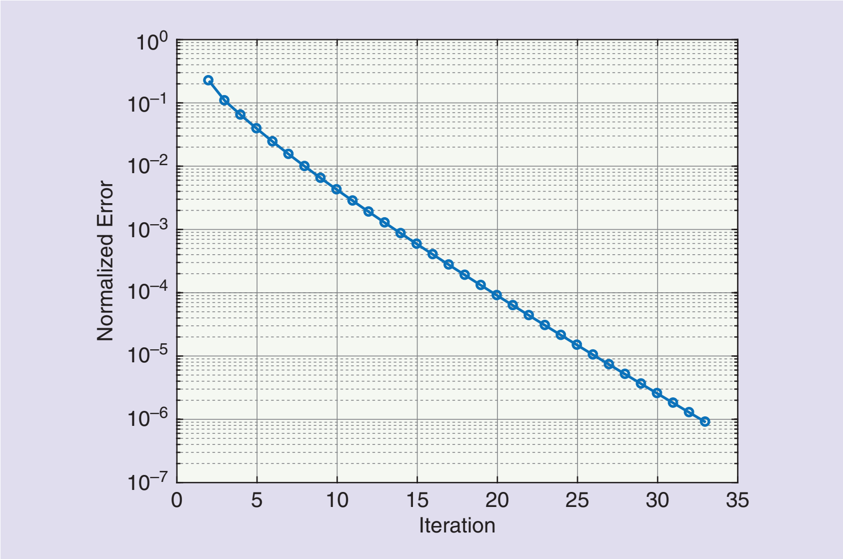

FIGURE 3. The normalized error of low-rank matrix completion with respect to the iteration count via gradient descent with the spectral initialization for a $10^4 \times 10^4$ matrix of rank-10 using about 5% observations.

Recall the SVD of $\boldsymbol{M}$ in (1), and denote $\boldsymbol{X}^\natural = \boldsymbol{U}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{Y}^\natural = \boldsymbol{V}\boldsymbol{\Sigma}^{1/2}$; this allows us to write the factorization as $\boldsymbol{M} = \boldsymbol{X}^\natural \boldsymbol{Y}^{\natural\top}$ and call $\boldsymbol{Z}^\natural = [\boldsymbol{X}^{\natural\top}, \boldsymbol{Y}^{\natural\top}]^\top \in \mathbb{R}^{(n_1+n_2)\times r}$ the ground truth. Since $\boldsymbol{Z}^\natural$ is only identifiable up to orthonormal transforms, let the optimal transform between the tth iterate $\boldsymbol{Z}_t = [\boldsymbol{X}_t^\top, \boldsymbol{Y}_t^\top]^\top \in \mathbb{R}^{(n_1+n_2)\times r}$ and $\boldsymbol{Z}^\natural$ as

$$\boldsymbol{H}_t := \operatorname*{argmin}_{\boldsymbol{R}\in\mathbb{R}^{r\times r}, \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}} \left\| \boldsymbol{Z}_t\boldsymbol{R} - \boldsymbol{Z}^\natural \right\|_{\mathrm{F}}.$$

Assume the condition number $\kappa := \sigma_1/\sigma_r$ of $\boldsymbol{M}$ is a bounded constant, then as long as

$$p \geq C_1 \frac{\mu^3 r^3 \log^3 n}{n}$$

for some sufficiently large constant $C_1 > 0$, with high probability, the iterates satisfy [8]

$$\| \boldsymbol{Z}_t\boldsymbol{H}_t - \boldsymbol{Z}^\natural \|_{\mathrm{F}} \leq C_2 \rho^t \mu r \frac{1}{\sqrt{np}} \| \boldsymbol{Z}^\natural \|_{\mathrm{F}},$$
$$\forall t \geq 0,$$

where $C_2 > 0, 0 < \rho < 1$ are some constants, provided that the step size $0 < \eta_t \equiv \eta \leq 2/(25\kappa\sigma_1)$. Hence, gradient descent converges at a geometric rate, as soon as the number of measurements is on the order of $\mu^3 r^3 n \log^3 n$, which scales linearly in n up to logarithmic factors. To reach ϵ-accuracy, i.e., $\| \boldsymbol{Z}_t\boldsymbol{H}_t - \boldsymbol{Z}^\natural \|_{\mathrm{F}} / \| \boldsymbol{Z}^\natural \|_{\mathrm{F}} \leq \epsilon$, gradient descent needs an order of $\log(1/\epsilon)$ iterations. The number of iterations is independent of the problem size and therefore the computational cost is much cheaper in conjunction with low cost per iteration.

Summary

Table 1 summarizes the figures-of-merit of the discussed algorithms using state-of-the-art theory.

Table 1. Figure-of-merits for low-rank matrix completion in terms of order-wise sample complexity and computational complexity.

	Sample Complexity	Computational Complexity
Information-theoretic lower bound	$\mu n r \log n$	NP-hard
Nuclear norm minimization	$\mu n r \log^2 n$	Polynomial time
Gradient descent with spectral initialization	$\mu^3 n r^3 \log^3 n$	Linear time

Numerical example

Let $\boldsymbol{M}$ be a rank-10 matrix of size $10^4 \times 10^4$ with about 5% of observed entries, i.e., $p = 0.05$, where $\boldsymbol{X}^\natural$ and $\boldsymbol{Y}^\natural$ are generated with i.i.d. standard Gaussian entries. We implement gradient descent with spectral initialization to recover $\boldsymbol{M}$. Figure 3 plots the normalized error $\| \boldsymbol{X}_t\boldsymbol{Y}_t^\top - \boldsymbol{M} \|_{\mathrm{F}} / \| \boldsymbol{M} \|_{\mathrm{F}}$ with respect to the iteration counts, which verifies the geometric convergence predicted by the theory. Indeed, the normalized error is below 10^{-5} within 30 iterations!

What we have learned

Under mild statistical models, low-rank matrix completion admits efficient algorithms with provable near-optimal performance guarantees, using both convex and nonconvex optimization techniques. The theory and algorithms discussed herein can be extended to recover matrices that are approximately low rank using noisy measurements. Low-rank matrix completion can be viewed as a special case of low-rank matrix estimation using an underdetermined set of linear equations. Other linear measurement patterns are also actively studied, motivated by applications such as sensor network localization, phase retrieval, quantum state tomography, and so on. Furthermore, low-rank matrix completion can be made robust even when many of the observations are corrupted by outliers of arbitrary magnitudes, known as the *sparse* and *low-rank decomposition problem* [9].

Low-rank structures are ubiquitous in modern data science problems and becoming increasingly popular as a modeling tool. Understanding the algorithmic and theoretical properties of estimation of low-rank structures is still an active area of research that will have a growing impact in future years. For a recent survey on low-rank matrix estimation, please see [10].

Acknowledgments

This work was supported by Air Force Office of Scientific Research under FA9550-15-1-0205, the Office of Naval Research under N00014-18-1-2142, and the National Science Foundation under CAREER ECCS-1818571 and CCF-1806154.

Author

Yuejie Chi (yuejiechi@cmu.edu) received her B.E. (Hon.) degree in electrical engineering from Tsinghua University, Beijing, China, in 2007 and her Ph.D. degree in electrical engineering from Princeton University, New Jersey, in 2012. She is currently an associate professor with the Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, Pennsylvania. Her research interests include statistical signal processing, machine learning, and large-scale optimization and their applications in data science, inverse problems, imaging, and sensing systems. She is a Senior Member of the IEEE.

References

[1] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Apr. 2009.

[2] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[3] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] D. Gross, “Recovering low-rank matrices from few coefficients in any basis,” *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, Mar. 2011.

[5] Y. Chen, “Incoherence-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2909–2923, 2015.

[6] R. H. Keshavan, A. Montanari, and S. Oh, “Matrix completion from a few entries,” *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.

[7] R. Sun and Z.-Q. Luo, “Guaranteed matrix completion via non-convex factorization,” *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6535–6579, 2016.

[8] C. Ma, K. Wang, Y. Chi, and Y. Chen, “Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution,” arXiv Preprint, arXiv:1711.10467, 2017.

[9] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *J. ACM*, vol. 58, no. 3, pp. 11, 2011.

[10] Y. Chen and Y. Chi, “Harnessing structures in big data via guaranteed low-rank matrix estimation,” *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 14–31, July 2018.