REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)
28-10-2021

2. REPORT TYPE
Final Report

3. DATES COVERED (From - To)
2-Jul-2018 - 1-Oct-2021

4. TITLE AND SUBTITLE
Final Report: Nonconvex Information Processing for Heterogeneous and Distributed Data

5a. CONTRACT NUMBER
W911NF-18-1-0303

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Princeton University
PO Box 36
87 Prospect Avenue, Second Floor
Princeton, NJ 08544 -2020

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
72851-CS.15

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Yuxin Chen
UU	UU	UU	UU		19b. TELEPHONE NUMBER 609-258-3996

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 26-Jan-2022

Agency Code: 21XD

Proposal Number: 72851CS

Agreement Number: W911NF-18-1-0303

INVESTIGATOR(S):

Name: Yuxin Chen
Email: yuxin.chen@princeton.edu
Phone Number: 6092583996
Principal: Y

Organization: **Princeton University**
Address: PO Box 36, Princeton, NJ 085442020
Country: USA
DUNS Number: 002484665
EIN: 210634501

Report Date: 01-Jan-2022
Date Received: 28-Oct-2021
Final Report for Period Beginning 02-Jul-2018 and Ending 01-Oct-2021
Title: Nonconvex Information Processing for Heterogeneous and Distributed Data
Begin Performance Period: 02-Jul-2018
End Performance Period: 01-Oct-2021
Report Term: 0-Other
Submitted By: Yuxin Chen
Email: yuxin.chen@princeton.edu
Phone: (609) 258-3996
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees:

STEM Participants:

Major Goals: The objective of this research program is to develop a comprehensive framework for nonconvex information processing with large-scale heterogeneous and distributed data. We aim to design efficient, robust and provably accurate algorithms for information processing using their natural nonconvex formulations without resorting to expensive convex relaxations. We will pursue a deeper understanding of geometric properties of nonconvex loss surfaces and optimization trajectories, with the emphasis on more realistic data models that can be heterogeneous and distributed. Novel combinations of insights and techniques from statistical signal processing, mathematical optimization, information theory, and high-dimensional statistics will be developed throughout the proposed research program to meet the research objectives.

Accomplishments: The PI and the co-PI have achieved the following significant results during this period.

1. We have developed a suite of nonconvex optimization algorithms for various nonconvex problems (including phase retrieval, matrix completion, tensor completion, tensor estimation, blind deconvolution, mixed matrix sensing).

2. We have demonstrated the unreasonable efficiency of nonconvex optimization algorithms even in the presence of random initialization, and have uncovered an implicit regularization phenomenon that underlies the success of nonconvex algorithms.

3. We have developed a suite of efficient uncertainty quantification schemes tailored to nonconvex optimization algorithms.

4. We have developed a suite of communication-efficient decentralized algorithm (e.g. NetworkDANE) that enables fast convergence in multi-agent distributed optimization settings.

5. We have investigated the effectiveness of spectral methods and subspace estimation methods in dealing with heterogeneous and heteroskedastic data.

Training Opportunities: Over the past three years, 10 PhD students and 3 postdocs have been involved in this project, which provide valuable opportunities for their research training.

Results Dissemination: PI Chen has given more than more than 60 talks in US universities and various workshops/conferences to disseminate the results; PI Chi has also given a number of invited talks and has been selected as the Goldsmith lecturer. PI Chen and co-PI Chi have finished an invited overview article on nonconvex low-rank factorization in IEEE Transactions on Signal Processing. PI Chen and co-PI Chi have just published a 240-page monograph "Spectral Methods for Data Science" in Foundations and Trends in Machine Learning. In addition, PI Chen and co-PI Chi have co-taught two tutorials on nonconvex optimization methods.

Honors and Awards: Yuxin Chen has received the 2020 ICCM Best paper award (gold medal).

Yuxin Chen has received the 2021 Princeton SEAS junior faculty award.

Yuxin Chen has received the 2020 ARO YIP award.

Yuxin Chen has been selected as a Finalist for the Best Paper Prize for Young Researchers in Continuous Optimization, 2019.

Yuejie Chi has been named the 2021 Goldsmith Lecturer by the IEEE Information Theory Society.

Yuejie Chi has won the Presidential Early Career Award for Scientists and Engineers (PECASE), 2019.

Yuejie Chi has been selected as the Inaugural recipient of IEEE Signal Processing Society Pierre-Simon Laplace Early Career Technical Achievement Award.

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: PD/PI
Participant: Yuxin Chen
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Pierre Bayle
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yanxi Chen
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yuling Yan
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:

National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Mengxin Yu
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Yuchen Zhou
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Qinghua Liu
Person Months Worked: 5.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Pengkun Yang
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Co PD/PI
Participant: Yuejie Chi
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Cong Ma
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Vincent Monardo
Person Months Worked: 3.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Maxime Ferreira Da Costa
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Shicong Cen
Person Months Worked: 10.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Harlee Lee
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Boyue Li
Person Months Worked: 1.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: The Annals of Statistics
Publication Identifier Type: DOI Publication Identifier: 10.1214/18-AOS1745
Volume: 47 Issue: 4 First Page #: 2204
Date Submitted: 6/7/20 12:00AM Date Published: 8/1/19 8:00AM
Publication Location:
Article Title: Spectral method and regularized MLE are both optimal for top-K ranking
Authors: Yuxin Chen, Jianqing Fan, Cong Ma, Kaizheng Wang
Keywords: Top-K ranking; pairwise comparisons; spectral method; regularized MLE; entrywise perturbation; leave-one-out analysis; reversible Markov chains
Abstract: This paper is concerned with the problem of top-K ranking from pairwise comparisons. Given a collection of n items and a few pairwise comparisons across them, one wishes to identify the set of K items that receive the highest ranks. To tackle this problem, we adopt the logistic parametric model—the Bradley–Terry–Luce model, where each item is assigned a latent preference score, and where the outcome of each pairwise comparison depends solely on the relative scores of the two items involved. Recent works have made significant progress toward characterizing the performance (e.g., the mean square error for estimating the scores) of several classical methods, including the spectral method and the maximum likelihood estimator (MLE). However, where they stand regarding top-K ranking remains unsettled. We demonstrate that under a natural random sampling model, the spectral method alone, or the regularized MLE alone, is minimax optimal in terms of the sample complexity.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Foundations of Computational Mathematics
Publication Identifier Type: DOI Publication Identifier: 10.1007/s10208-019-09429-9
Volume: 20 Issue: 3 First Page #: 451
Date Submitted: 6/7/20 12:00AM Date Published: 8/1/19 4:00AM
Publication Location:
Article Title: Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval, Matrix Completion, and Blind Deconvolution
Authors: Cong Ma, Kaizheng Wang, Yuejie Chi, Yuxin Chen
Keywords: Nonconvex optimization; Gradient descent; Leave-one-out analysis; Phase retrieval; Matrix completion; Blind deconvolution
Abstract: Recent years have seen a flurry of activities in designing provably efficient nonconvex procedures for solving statistical estimation problems. Due to the highly nonconvex nature of the empirical loss, state-of-the-art procedures often require proper regularization (e.g., trimming, regularized cost, projection) in order to guarantee fast convergence. For vanilla procedures such as gradient descent, however, prior theory either recommends highly conservative learning rates to avoid overshooting, or completely lacks performance guarantees. This paper uncovers a striking phenomenon in nonconvex optimization: even in the absence of explicit regularization, gradient descent enforces proper regularization implicitly under various statistical models. In fact, gradient descent follows a trajectory staying within a basin that enjoys nice geometry, consisting of points incoherent with the sampling mechanism.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 26-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: SIAM Journal on Optimization
Publication Identifier Type: DOI Publication Identifier: 10.1137/19M1290000
Volume: 30 Issue: 4 First Page #: 3098
Date Submitted: 5/29/21 12:00AM Date Published: 10/1/20 4:00AM
Publication Location:
Article Title: Noisy Matrix Completion: Understanding Statistical Guarantees for Convex Relaxation via Nonconvex Optimization
Authors: Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, Yuling Yan
Keywords: matrix completion, minimaxity, stability, convex relaxation, nonconvex optimization, Burer– Monteiro approach
Abstract: This paper studies noisy low-rank matrix completion: given partial and noisy entries of a large low-rank matrix, the goal is to estimate the underlying matrix faithfully and efficiently. Arguably one of the most popular paradigms to tackle this problem is convex relaxation, which achieves remarkable efficacy in practice. However, the theoretical support of this approach is still far from optimal in the noisy setting, falling short of explaining its empirical success. We make progress towards demystifying the practical efficacy of convex relaxation vis-à-vis random noise. When the rank and the condition number of the unknown matrix are bounded by a constant, we demonstrate that the convex programming approach achieves near-optimal estimation errors --- in terms of the Euclidean loss, the entrywise loss, and the spectral norm loss --- for a wide range of noise levels. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct alg
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Proceedings of the National Academy of Sciences
Publication Identifier Type: DOI Publication Identifier: 10.1073/pnas.1910053116
Volume: 116 Issue: 46 First Page #: 22931
Date Submitted: 5/29/21 12:00AM Date Published: 11/1/19 4:00AM
Publication Location:
Article Title: Inference and uncertainty quantification for noisy matrix completion
Authors: Yuxin Chen, Jianqing Fan, Cong Ma, Yuling Yan
Keywords: matrix completion, statistical inference, confidence intervals, uncertainty quantification, convex relaxation, nonconvex optimization
Abstract: Noisy matrix completion aims at estimating a low-rank matrix given only partial and corrupted entries. Despite substantial progress in designing efficient estimation algorithms, it remains largely unclear how to assess the uncertainty of the obtained estimates and how to perform statistical inference on the unknown matrix (e.g. constructing a valid and short confidence interval for an unseen entry). This paper takes a step towards inference and uncertainty quantification for noisy matrix completion. We develop a simple procedure to compensate for the bias of the widely used convex and nonconvex estimators. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals / regions for, say, the missing entries and the low-rank factors. Our inferential procedures do not rely on sample splitting, thus avoiding unnecessary loss of data efficiency. As a byproduct, we obtain a sharp char
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 26-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TSP.2019.2937282
Volume: 67 Issue: 20 First Page #: 5239
Date Submitted: 5/29/21 12:00AM Date Published: 10/1/19 4:00AM
Publication Location:
Article Title: Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview
Authors: Yuejie Chi, Yue M. Lu, Yuxin Chen
Keywords: First-order methods, landscape analysis, matrix factorization, nonconvex optimization, statistics
Abstract: Substantial progress has been made recently on developing provably accurate and efficient algorithms for low-rank matrix factorization via nonconvex optimization. While conventional wisdom often takes a dim view of nonconvex optimization algorithms due to their susceptibility to spurious local minima, simple iterative methods such as gradient descent have been remarkably successful in practice. The theoretical footings, however, had been largely lacking until recently. In this tutorial-style overview, we highlight the important role of statistical models in enabling efficient nonconvex optimization with performance guarantees. We review two contrasting approaches: (1) two-stage algorithms, which consist of a tailored initialization step followed by successive refinement; and (2) global landscape analysis and initialization-free algorithms. Several canonical matrix factorization problems are discussed, including but not limited to matrix sensing, phase retrieval, matrix completion, blin
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: The Annals of Statistics
Publication Identifier Type: DOI Publication Identifier: 10.1214/20-AOS1986
Volume: 49 Issue: 2 First Page #:
Date Submitted: 5/29/21 12:00AM Date Published: 4/1/21 4:00AM
Publication Location:
Article Title: Subspace estimation from unbalanced and incomplete data matrices: $L_{2,\infty}$ statistical guarantees
Authors: Changxiao Cai, Gen Li, Yuejie Chi, H. Vincent Poor, Yuxin Chen
Keywords: spectral method, principal component analysis with missing data, tensor completion, covariance estimation, spectral clustering, leave-one-out analysis
Abstract: This paper is concerned with estimating the column space of an unknown low-rank matrix, given noisy and partial observations of its entries. There is no shortage of scenarios where the observations — while being too noisy to support faithful recovery of the entire matrix — still convey sufficient information to enable reliable estimation of the column space of interest. This is particularly evident and crucial for the highly unbalanced case where the column dimension d2 far exceeds the row dimension d1, which is the focal point of the current paper. We investigate an efficient spectral method, which operates upon the sample Gram matrix with diagonal deletion. While this algorithmic idea has been studied before, we establish new statistical guarantees for this method in terms of both ?2 and ?2,? estimation accuracy, which improve upon prior results if d2 is substantially larger than d1. To illustrate the effectiveness of our findings, we derive matching minimax lower bound
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 26-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Information Theory
Publication Identifier Type: DOI Publication Identifier: 10.1109/TIT.2021.3050427
Volume: 67 Issue: 3 First Page #: 1928
Date Submitted: 5/29/21 12:00AM Date Published: 3/1/21 5:00AM
Publication Location:
Article Title: Nonconvex Matrix Factorization From Rank-One Measurements
Authors: Yuanxin Li, Cong Ma, Yuxin Chen, Yuejie Chi
Keywords: matrix factorization, rank-one measurements, gradient descent, nonconvex optimization
Abstract: We consider the problem of recovering low-rank matrices from random rank-one measurements, which spans numerous applications including covariance sketching, phase retrieval, quantum state tomography, and learning shallow polynomial neural networks, among others. Our approach is to directly estimate the low-rank factor by minimizing a nonconvex quadratic loss function via vanilla gradient descent, following a tailored spectral initialization. When the true rank is small, this algorithm is guaranteed to converge to the ground truth (up to global ambiguity) with near-optimal sample complexity and computational complexity. To the best of our knowledge, this is the first guarantee that achieves near-optimality in both metrics. In particular, the key enabler of near-optimal computational guarantees is an implicit regularization phenomenon: without explicit regularization, both spectral initialization and the gradient descent iterates automatically stay within a region incoherent with the mea
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: The Annals of Statistics
Publication Identifier Type: DOI Publication Identifier: 10.1214/20-AOS1963
Volume: 49 Issue: 1 First Page #:
Date Submitted: 5/29/21 12:00AM Date Published: 2/1/21 5:00AM
Publication Location:
Article Title: Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices
Authors: Yuxin Chen, Chen Cheng, Jianqing Fan
Keywords: asymmetric matrices, eigenvalue perturbation, entrywise eigenvector perturbation, linear forms of eigenvectors, heteroscedasticity
Abstract: This paper is concerned with the interplay between statistical asymmetry and spectral methods. Suppose we are interested in estimating a rank-1 and symmetric matrix, yet only a randomly perturbed version M is observed. The noise matrix is composed of zero-mean independent (but not necessarily homoscedastic) entries and is, therefore, not symmetric in general. This might arise, for example, when we have two independent samples for each entry of Mstar and arrange them into an {\em asymmetric} data matrix M. The aim is to estimate the leading eigenvalue and eigenvector of Mstar. We demonstrate that the leading eigenvalue of the data matrix M can be sqrt{n} times more accurate --- up to some log factor --- than its (unadjusted) leading singular value in eigenvalue estimation. Further, the perturbation of any linear form of the leading eigenvector of M --- say, entrywise eigenvector perturbation --- is provably well-controlled. This eigen-decomposition approach is fully adaptive to he
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Mathematical Programming
Publication Identifier Type: DOI Publication Identifier: 10.1007/s10107-019-01363-6
Volume: 176 Issue: First Page #: 5
Date Submitted: 5/29/21 12:00AM Date Published: 2/1/19 5:00AM
Publication Location:
Article Title: Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval
Authors: Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma
Keywords: gradient descent, random initialization, phase retrieval
Abstract: This paper considers the problem of solving systems of quadratic equations, namely, recovering an object of interest. This problem, also dubbed as phase retrieval, spans multiple domains including physical sciences and machine learning. We investigate the efficacy of gradient descent (or Wirtinger flow) designed for the nonconvex least squares problem. We prove that under Gaussian designs, gradient descent — when randomly initialized— yields an epsilon-accurate solution in a logarithmic number of iterations given nearly minimal samples, thus achieving near-optimal computational and sample complexities at once. This provides the first global convergence guarantee concerning vanilla gradient descent for phase retrieval, without the need of (i) carefully-designed initialization, (ii) sample splitting, or (iii) sophisticated saddle-point escaping schemes.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Geoscience and Remote Sensing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TGRS.2020.3020810
Volume: 59 Issue: 4 First Page #: 3338
Date Submitted: 6/12/21 12:00AM Date Published: 4/1/21 4:00AM
Publication Location:
Article Title: Blind Hyperspectral Unmixing Based on Graph Total Variation Regularization
Authors: Jing Qin, Harlin Lee, Jocelyn T. Chi, Lucas Drumetz, Jocelyn Chanussot, Yifei Lou, Andrea L. Bertozzi
Keywords: ADMM; hyperspectral imaging
Abstract: Remote sensing data from hyperspectral cameras suffer from limited spatial resolution, in which a single pixel of a hyperspectral image may contain information from several materials in the field of view. Blind hyperspectral image unmixing is the process of identifying the pure spectra of individual materials (i.e., endmembers) and their proportions (i.e., abundances) at each pixel. In this article, we propose a novel blind hyperspectral unmixing model based on the graph total variation (gTV) regularization, which can be solved efficiently by the alternating direction method of multipliers (ADMM). To further alleviate the computational cost, we apply the Nyström method to approximate a fully connected graph by a small subset of sampled points. Furthermore, we adopt the Merriman–Bence– Osher (MBO) scheme to solve the gTV-involved subproblem in ADMM by decomposing a gray-scale image into a bitwise form.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 26-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TSP.2021.3071560
Volume: 69 Issue: First Page #: 2396
Date Submitted: 6/12/21 12:00AM Date Published:
Publication Location:
Article Title: Low-Rank Matrix Recovery With Scaled Subgradient Methods: Fast and Robust Convergence Without the Condition Number
Authors: Tian Tong, Cong Ma, Yuejie Chi
Keywords: Low-rank matrix recovery, nonsmooth and nonconvex optimization, residual sum of absolute errors, scaled subgradient methods
Abstract: Many problems in data science can be treated as estimating a low-rank matrix from highly incomplete, sometimes even corrupted, observations. One popular approach is to resort to matrix factorization, where the low-rank matrix factors are optimized via first-order methods over a smooth loss function, such as the residual sum of squares.While tremendous progress has been made in recent years, the natural smooth formulation suffers from two sources of ill-conditioning, where the iteration complexity of gradient descent scales poorly both with the dimension as well as the condition number of the low-rank matrix.Moreover, the smooth formulation is not robust to corruptions. In this paper, we propose scaled subgradient methods to minimize a family of nonsmooth and nonconvex formulations—in particular, the residual sum of absolute errors—which is guaranteed to converge at a fast rate that is almost dimension-free and independent of the condition number, even in the presence of corruptions. We
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TSP.2021.3051425
Volume: 69 Issue: First Page #: 867
Date Submitted: 6/12/21 12:00AM Date Published:
Publication Location:
Article Title: Beyond Procrustes: Balancing-Free Gradient Descent for Asymmetric Low-Rank Matrix Sensing
Authors: Cong Ma, Yuanxin Li, Yuejie Chi
Keywords: Asymmetric low-rank matrix sensing, nonconvex optimization, gradient descent
Abstract: Low-rank matrix estimation plays a central role in various applications across science and engineering. Recently, nonconvex formulations based on matrix factorization are provably solved by simple gradient descent algorithms with strong computational and statistical guarantees. However, when the low-rank matrices are asymmetric, existing approaches rely on adding a regularization term to balance the scale of the two matrix factors which in practice can be removed safely without hurting the performance when initialized via the spectral method. In this paper, we provide a theoretical justification to this for the matrix sensing problem, which aims to recover a low-rank matrix from a small number of linear measurements. As long as the measurement ensemble satisfies the restricted isometry property, gradient descent—in conjunction with spectral initialization—converges linearly without the need of explicitly promoting balancedness of the factors.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 26-Jan-2022

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Information Theory
Publication Identifier Type: DOI Publication Identifier: 10.1109/TIT.2021.3111828
Volume: 67 Issue: 11 First Page #: 7380
Date Submitted: 10/28/21 12:00AM Date Published: 11/1/21 7:00AM
Publication Location:
Article Title: Tackling Small Eigen-Gaps: Fine-Grained Eigenvector Estimation and Inference Under Heteroscedastic Noise
Authors: Chen Cheng, Yuting Wei, Yuxin Chen
Keywords: Eigen-gap, linear form of eigenvectors, confidence interval, uncertainty quantification, heteroscedasticity
Abstract: This paper aims to address two fundamental challenges arising in eigenvector estimation and inference for a low-rank matrix from noisy observations: 1) how to estimate an unknown eigenvector when the eigen-gap (i.e. the spacing between the associated eigenvalue and the rest of the spectrum) is particularly small; 2) how to perform estimation and inference on linear functionals of an eigenvector—a sort of “fine-grained” statistical reasoning that goes far beyond the usual 2 analysis. We investigate how to address these challenges in a setting where the unknown n × n matrix is symmetric and the additive noise matrix contains independent (and non-symmetric) entries. Based on eigen-decomposition of the asymmetric data matrix, we propose estimation and uncertainty quantification procedures for an unknown eigenvector, which further allow us to reason about linear functionals of an unknown eigenvector.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Foundations and Trends® in Machine Learning
Publication Identifier Type: DOI Publication Identifier: 10.1561/2200000079
Volume: 14 Issue: 5 First Page #: 566
Date Submitted: 10/28/21 12:00AM Date Published:
Publication Location:
Article Title: Spectral Methods for Data Science: A Statistical Perspective
Authors: Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma
Keywords: spectral methods; leave-one-out analysis
Abstract: This paper aims to address two fundamental challenges arising in eigenvector estimation and inference for a low-rank matrix from noisy observations: 1) how to estimate an unknown eigenvector when the eigen-gap (i.e. the spacing between the associated eigenvalue and the rest of the spectrum) is particularly small; 2) how to perform estimation and inference on linear functionals of an eigenvector—a sort of “fine-grained” statistical reasoning that goes far beyond the usual 2 analysis. We investigate how to address these challenges in a setting where the unknown n × n matrix is symmetric and the additive noise matrix contains independent (and non-symmetric) entries. Based on eigen-decomposition of the asymmetric data matrix, we propose estimation and uncertainty quantification procedures for an unknown eigenvector, which further allow us to reason about linear functionals of an unknown eigenvector.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Partners

,

I certify that the information in the report is complete and accurate:
Signature: Yuxin Chen
Signature Date: 10/28/21 2:09AM

ARO Annual Report (August 2020 - July 2021)
Nonconvex Information Processing for Heterogeneous and Distributed Data

PIs: Yuxin Chen (Princeton University) and Yuejie Chi (Carnegie Mellon University)

October 27, 2021

Novel theory and algorithms in scalable nonconvex information processing. The importance of succinct and informative latent structures cannot be overstated in large-scale information processing. Many high-dimensional data possess low-complexity structures whose intrinsic degrees-of-freedom are much smaller than the ambient data dimension — a key enabler of information recovery from highly incomplete datasets. However, we are often faced with fundamental computational and statistical challenges when extracting low-complexity latent structures in large-scale information processing tasks. For instance, the data acquisition mechanisms are often highly nonlinear and complex mappings of the latent structures, making most classical statistical estimators highly nonconvex and intractable in general. Also, existing theory fails to predict when and why simple first-order optimization methods are effective in solving nonconvex problems in practice.

In this ARO project, we have made substantial progress towards developing: (1) new theory that better explains / predicts practice and that provides more insights for practitioners to keep the momentum, and (2) new algorithms that possess both statistical and computational guarantees when accommodating the ever-increasing volume of data. In particular, several highlights of our achievements are:

- *Implicit regularization in nonconvex information processing.* In order to enable efficient nonconvex information processing, many prior theoretical works developed complicated procedures based on sophisticated regularization schemes. Such regularization schemes rely heavily on idealistic statistical models and introduce extra tuning parameters that need to be carefully tuned. These could be quite stringent and fragile when handling realistic scenarios, significantly hampering their practical applicability. By contrast, our works developed a suite of regularization-free nonconvex optimization methods that rely — in a fairly insensitive manner — on minimal tuning parameters. We demonstrate the unreasonable effectiveness of unregularized first-order methods in several practically important problems, including matrix completion, blind deconvolution, phase retrieval, covariance sketching, and tensor completion. In all of these problems, we discover that the algorithms implicitly regularize the running

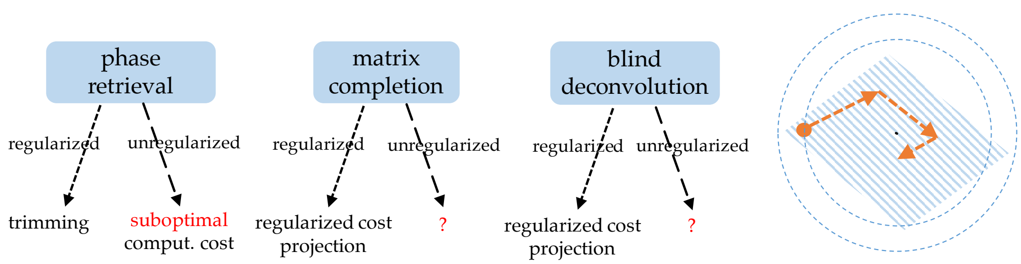

Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

iterates in a way that favors fast convergence. This intriguing observations have important implications on real applications, suggesting that one can safely resort to model-agnostic optimization-based methods (which are certainly more robust in the presence of model mismatch) without compromising any statistical accuracy.

- *Random initialization in nonconvex information processing.* Another issue that often confuses researchers is the necessity of of carefully-designed initialization when solving nonconvex problems. In order to guarantee fast convergence, prior theory typically recommend a "warm start". On the contrary, practitioners usually like to initialize the algorithms randomly — an approach that is significantly more robust against model mismatch. However, the theoretical performance of randomly initialized gradient methods is poorly understood for nonconvex problems. The only result one can find in prior literature says that randomly initialized gradient methods converge almost surely to the local optimizers, without any guarantees on the convergence rates; in fact, this cannot even ensure polynomial-time convergence of the algorithm. Our works have obtained a theoretical breakthrough in understanding randomly initialized gradient methods. Focusing on the phase retrieval problem (or solving quadratic systems of equations), we show that randomly initialized gradient methods achieve the desired statistical accuracy in a small number of iterations; more specifically, the algorithm provably attains ε accuracy within $O(\log n + \log \frac{1}{\varepsilon})$ iterations, thus achieving the fastest possible convergence. The convergence guarantee we derive is at least $\mathsf{poly}(n)$ times faster than existing sophisticated saddle-escaping algorithms (where n is the dimension of the signal, which is a very large number), even though the gradient algorithm we analyze is completely model-agnostic. Furthermore, the algorithm also achieves minimax-optimal statistical accuracy in the presence of noise, thus indicating that a simple optimization-based method achieves optimal statistical accuracy and computational efficiency all at once.

- *Fast and statistically optimal tensor completion via nonconvex optimization.* In this work, we study a problem of broad practical interest, namely, the reconstruction of a large-dimensional low-rank tensor from highly incomplete and randomly corrupted observations of its entries. While a number of papers have been dedicated to this tensor completion problem, prior algorithms either are computationally too expensive for large-scale applications, or come with sub-optimal statistical performance. Motivated by this, we propose a fast two-stage nonconvex algorithm — gradient method following a rough initialization — that achieves the best of both worlds: optimal statistical accuracy and computational effiency. Specifically, the proposed algorithm provably completes the tensor and retrieves all low-rank factors within nearly linear time, while at the same time enjoying near-optimal statistical guarantees (i.e. minimal sample complexity and optimal estimation accuracy). The insights conveyed through our analysis of nonconvex optimization might have implications for a broader family of tensor reconstruction problems beyond tensor completion.

- *Uncertainty quantification for nonconvex tensor completion in the presence of heteroskedastic data.* Moving from estimation to uncertainty quantification (an important step towards trustworthy decision making), we study the distribution and uncertainty of nonconvex optimization for noisy tensor completion. Focusing on the above-mentioned two-stage estimation algorithm, we characterize the distribution of our nonconvex estimator down to fine scales. This distributional theory in turn allows one to construct valid and short confidence intervals for both the unseen tensor entries and the unknown tensor factors. The proposed inferential procedure enjoys several important features: (1) it is fully adaptive to noise heteroscedasticity, and (2) it is data-driven and automatically adapts to unknown noise distributions. Furthermore, our findings unveil the statistical optimality of nonconvex tensor completion: it attains un-improvable Euclidean accuracy—including both the rates and the pre-constants—when estimating both the unknown tensor and the underlying tensor factors.

- *Fast global convergence of nonconvex policy optimization.* Another approach that lies at the heart of recent RL advances is the class of policy gradient (PG) methods and their variants, which aim to optimize (parameterized) policies via gradient-type methods. As an important and widely used extension of PG methods, natural policy gradient methods propose to employ natural policy gradients as search directions, in order to achieve faster convergence than the update rules based on policy gradients. Despite the enormous empirical success, the theoretical underpinnings of NPG methods

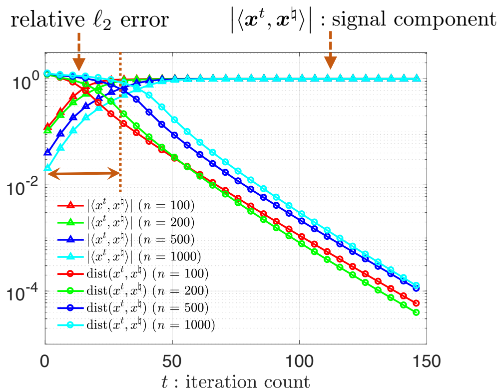

Figure 2: Numerically convergence of randomly initialized gradient methods.

have been severely limited even until recently, primarily due to the intrinsic non-concavity underlying the value maximization problem of interest. Our recent work pursues in-depth non-asymptotic understanding about the efficacy of entropy-regularized NPG methods — a paradigm that has been found remarkably effective in practice. We prove that the regularized NPG method converges linearly (or even quadratically once it enters a local region around the optimal policy), and is provably stable vis-à-vis inexactness of policy gradients. Our convergence guarantees accommodate a wide range of learning rates of practical value, and shed light upon the role of entropy regularization in enabling fast convergence for policy optimization.

- *Communication-efficient distributed optimization in resource-constrained network environments.* There is growing interest in large-scale machine learning and optimization over decentralized networks, e.g. in the context of multi-agent learning and federated learning. Due to the imminent need to alleviate the communication burden, the investigation of communication-efficient distributed optimization algorithms has flourished in recent years. In this work, we study distributed optimization over networks, where each agent is only allowed to aggregate information from its neighbors over a network (namely, no centralized coordination is present). By properly adjusting the global gradient estimate via local averaging in conjunction with proper correction, we develop a communication-efficient approximate Newton-type method, called Network-DANE, which generalizes DANE to accommodate decentralized scenarios. We establish fast convergence of our algorithms for various scenarios, which shed light on the impacts of data homogeneity, network connectivity, and local averaging upon the rate of convergence. Our work suggests that by performing a judiciously chosen amount of local communication and computation per iteration, the overall efficiency can be substantially improved.

- *Accelerating ill-conditioned nonconvex low-rank estimation.* Despite nonconvexity, recent literatures have shown that these simple nonconvex algorithms achieve linear convergence when initialized properly for a growing number of problems of interest. However, existing approaches can still be computationally expensive especially for ill-conditioned matrices: the convergence rate of gradient descent depends linearly on the condition number of the low-rank matrix, while the per-iteration cost of alternating minimization is often prohibitive for large matrices. To address this issue, we propose a competitive algorithmic approach dubbed Scaled Gradient Descent (ScaledGD) which can be viewed as preconditioned or diagonally-scaled gradient descent, where the preconditioners are adaptive and iteration-varying with a minimal computational overhead. With tailored variants for low-rank matrix sensing, robust principal component analysis and matrix completion, we theoretically show that ScaledGD achieves the best of both worlds: it converges linearly at a rate independent of the condition number of the low-rank matrix similar as alternating minimization, while maintaining the low per-iteration cost of gradient descent. To the best of our knowledge, ScaledGD is the first algorithm that provably has such properties over a wide range of low-rank matrix estimation tasks.

List of publications:

1. "Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data", Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, accepted to *Annals of Statistics*, 2021.

2. "Communication-efficient distributed optimization in networks with gradient tracking and variance reduction", B. Li, S. Cen, Y. Chen, Y. Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

3. "Accelerating Ill-Conditioned Low-Rank Matrix Estimation via Scaled Gradient Descent", T. Tong, C. Ma, and Y. Chi, *Journal of Machine Learning Research*, accepted, 2021.

4. "Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number", T. Tong, C. Ma, and Y. Chi, *IEEE Transactions on Signal Processing*, vol. 69, pp. 2396-2409, 2021.

5. "Noisy matrix completion: understanding statistical guarantees for convex relaxation via nonconvex optimization", Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *SIAM Journal on Optimization*, vol. 30, no. 4, pp. 3098–3121, 2020.

6. "Beyond Procrustes: Balancing-Free Gradient Descent for Asymmetric Low-Rank Matrix Sensing," C. Ma, Y. Li and Y. Chi, *IEEE Transactions on Signal Processing*, vol. 69, pp. 867-877, 2021.

7. "Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees", C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.

8. "Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices", Y. Chen, C. Cheng, J. Fan, accepted to the *Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021.

9. "Nonconvex low-rank tensor completion from noisy data", C. Cai, G. Li, H. V. Poor, Y. Chen, accepted to *Operations Research*, 2020.

10. "Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval", Y. Chen, Y. Chi, J. Fan, C. Ma, *Mathematical Programming*, vol. 176, no. 1-2, pp. 5-37, July 2019

11. "Implicit regularization in nonconvex statistical estimation: gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution", C. Ma, K. Wang, Y. Chi, Y. Chen, *Foundations of Computational Mathematics*, vol. 20, no. 3, pp. 451-632, June 2020

12. "Nonconvex optimization meets low-rank matrix factorization: an overview", Y. Chi, Y. Lu, Y. Chen, *IEEE Trans. Signal Processing*, vol. 47, no. 4, pp. 2204-2235, 2019 **(invited overview article)**

13. "Spectral method and regularized MLE are both optimal for top-K ranking", Y. Chen, J. Fan, C. Ma, K. Wang, *Annals of Statistics*, vol. 47, no. 4, pp. 2204-2235, August 2019

14. "Fast global convergence of natural policy gradient methods with entropy regularization", S. Cen, C. Cheng, Y. Chen, Y. Wei, Y. Chi, accepted to *Operations Research*, 2021.

15. "Softmax policy gradient methods can take exponential time to converge", G. Li, Y. Wei, Y. Chi, Y. Gu, Y. Chen, accepted to *Conference on Learning Theory (COLT)*, 2021.

16. "Inference and uncertainty quantification for noisy matrix completion", Y. Chen, J. Fan, C. Ma, Y. Yan, *Proceedings of the National Academy of Sciences (PNAS)*, vol. 116, no. 46, pp. 22931-22937, Nov. 2019

17. "Nonconvex low-rank symmetric tensor completion from noisy data", C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.

18. "Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality", C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.

19. "Spectral Methods for Data Science: A Statistical Perspective", *Foundations and Trends in Machine Learning*, Y. Chen, Y. Chi, J. Fan, C. Ma, vol. 14, no. 5, pp. 566-806, 2021.

20. "Learning Latent Features with Pairwise Penalties in Matrix Completion", K. Ji, J. Tan, Y. Chi and J. Xu, in revision to *IEEE Trans. on Signal Processing*, Feb. 2019.

21. "Guaranteed Recovery of One-Hidden-Layer Neural Networks via Cross Entropy", H. Fu, Y. Chi and Y. Liang, submitted to *IEEE Trans. on Signal Processing*, Jan. 2019.

22. "Median-Truncated Gradient Descent: A Robust and Scalable Approach for Nonconvex Signal Estimation", Y. Chi, Y. Li, H. Zhang, and Y. Liang, *Compressed Sensing and Its Applications*, Springer International Publishing, 2019+.

23. "Analytical Convergence Regions of Accelerated First-Order Methods in Nonconvex Optimization under Regularity Condition", H. Xiong, Y. Chi, B. Hu and W. Zhang, *Automatica*, provisionally accepted.

24. "Nonconvex Low-Rank Matrix Recovery with Arbitrary Outliers via Median-Truncated Gradient Descent", Y. Li, Y. Chi, H. Zhang, and Y. Liang, *Information and Inference: A Journal of the IMA*, in press.

25. "Nonconvex matrix factorization from rank-one measurements", Y. Li, C. Ma, Y. Chen, and Y. Chi, International Conference on *Artificial Intelligence and Statistics*, 2019

26. V. Monardo and Y. Chi, "On the Sensitivity of Spectral Initialization for Noisy Phase Retrieval", in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

27. V. Monardo, Y. Li and Y. Chi, "Solving Quadratic Equations via Amplitude-Based Nonconvex Optimization", in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

Invited tutorial

1. Y. Chen, Y. Chi, and C. Ma, "Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods," European Signal Processing Conference (EUSIPCO) 2020.

2. Y. Chen and Y. Chi, "Taming Nonconvexity in Information Science," *Information Theory Workshop (ITW) 2018.*