



AFRL-AFOSR-VA-TR-2023-0311

Taming Nonconvexity in Solving High-Dimensional Nonlinear Systems of Equations

Yuxin Chen
TRUSTEES OF PRINCETON UNIVERSITY
1 NASSAU HALL
PRINCETON, NJ,
US

03/20/2023
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20230320		2. REPORT TYPE Final		3. DATES COVERED	
				START DATE 20190101	END DATE 20221231
4. TITLE AND SUBTITLE Taming Nonconvexity in Solving High-Dimensional Nonlinear Systems of Equations					
5a. CONTRACT NUMBER		5b. GRANT NUMBER FA9550-19-1-0030		5c. PROGRAM ELEMENT NUMBER 61102F	
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER	
6. AUTHOR(S) Yuxin Chen					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) TRUSTEES OF PRINCETON UNIVERSITY 1 NASSAU HALL PRINCETON, NJ US					8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR RTA2	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-VA-TR-2023-0311
12. DISTRIBUTION/AVAILABILITY STATEMENT A Distribution Unlimited: PB Public Release					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU		18. NUMBER OF PAGES 7
19a. NAME OF RESPONSIBLE PERSON WARREN ADAMS					19b. PHONE NUMBER (Include area code) 00000000

Standard Form 298 (Rev. 5/2020)
Prescribed by ANSI Std. Z39.18

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

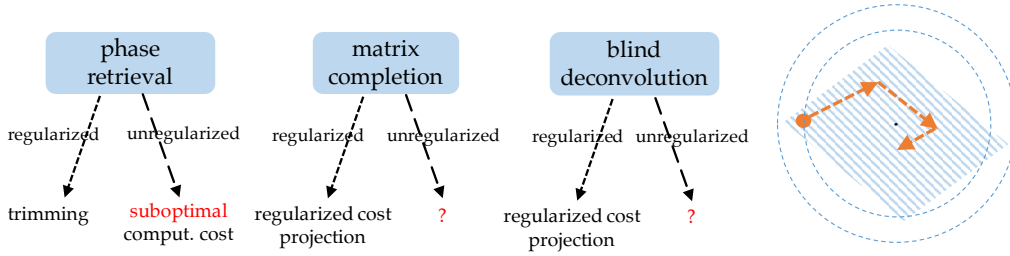


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

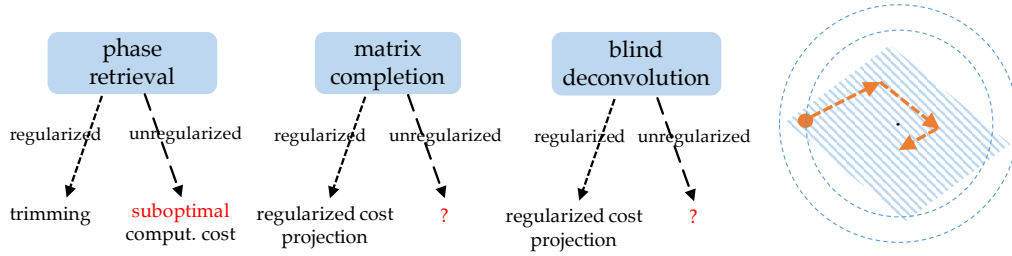


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

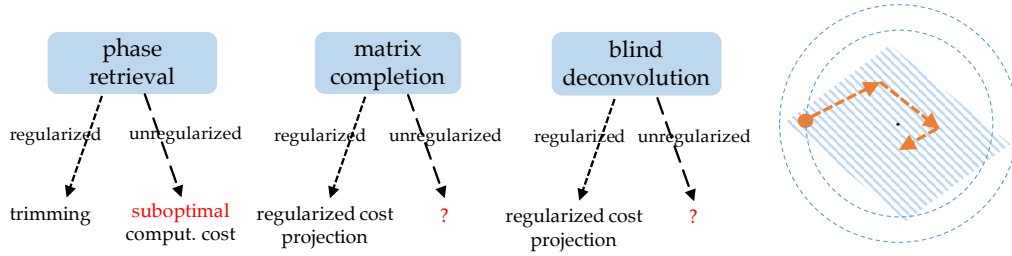


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

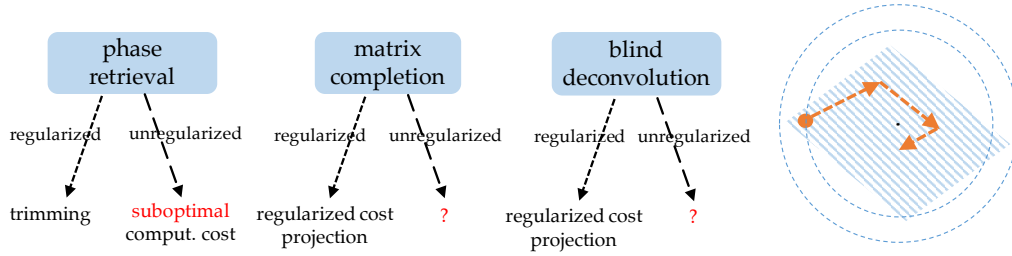


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

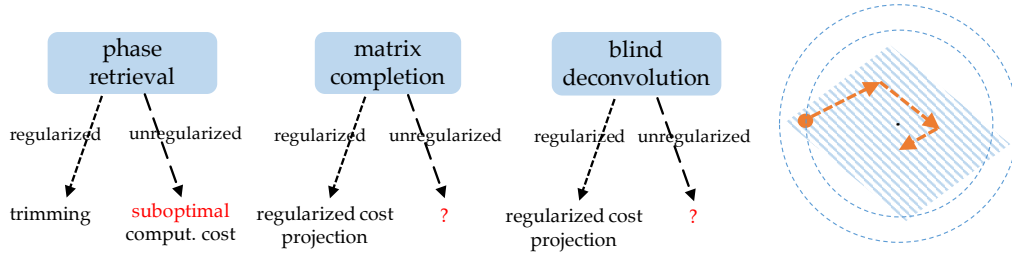


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

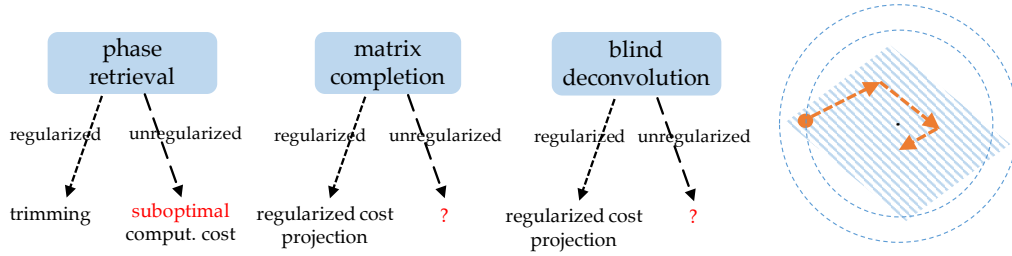


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

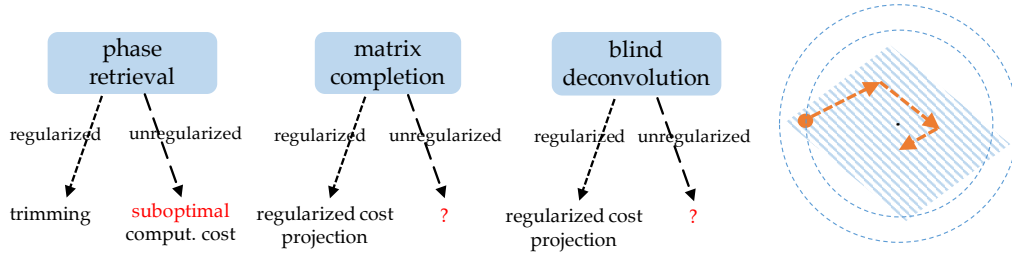


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model,” *Neural Information Processing Systems (NeurIPS)*, December 2020.

14. Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen, “Sample Complexity of Asynchronous Q-Learning: Sharper Analysis and Variance Reduction,” *Neural Information Processing Systems (NeurIPS)*, December 2020.
15. “The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square”, P. Sur, Y. Chen, E. Candes, *Probability Theory and Related Fields*, vol. 175, no. 1-2, pp. 487-558, October 2019
16. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.
17. “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
18. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
19. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
20. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
21. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
22. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
23. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
24. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

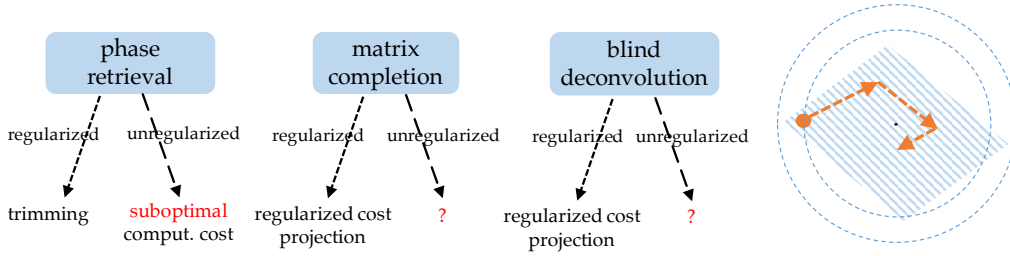


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

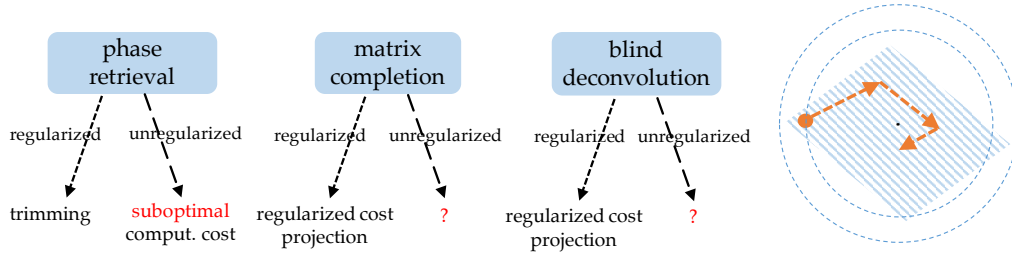


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

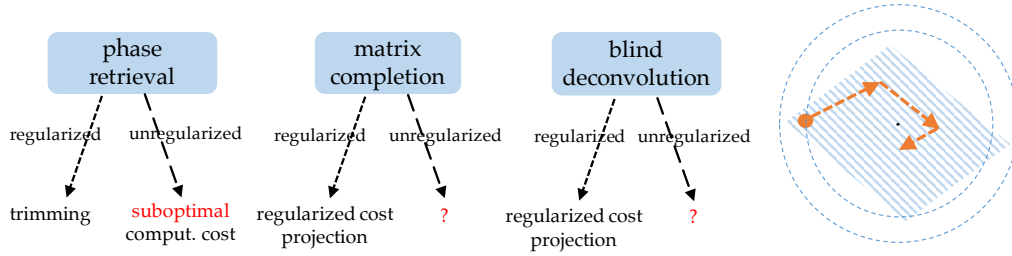


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

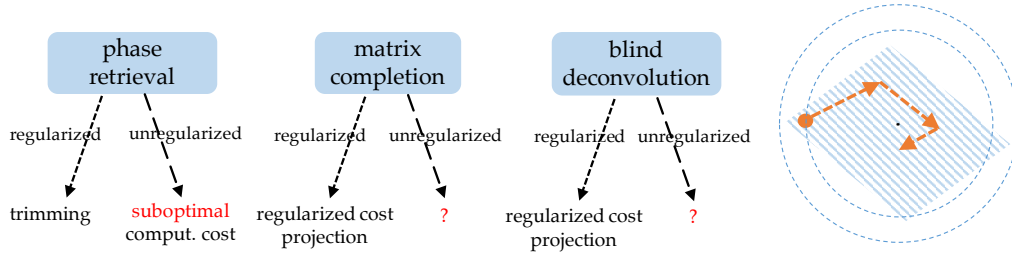


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

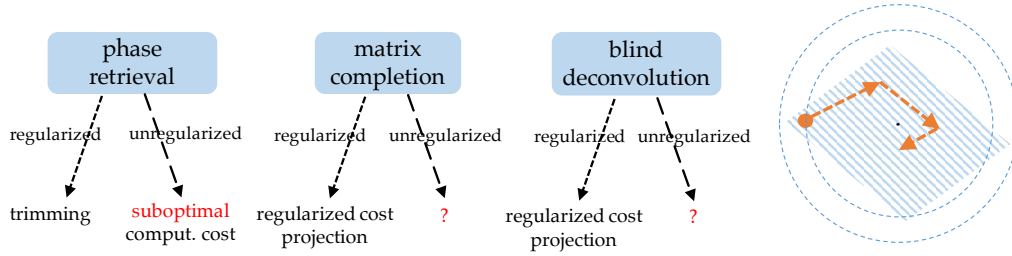


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

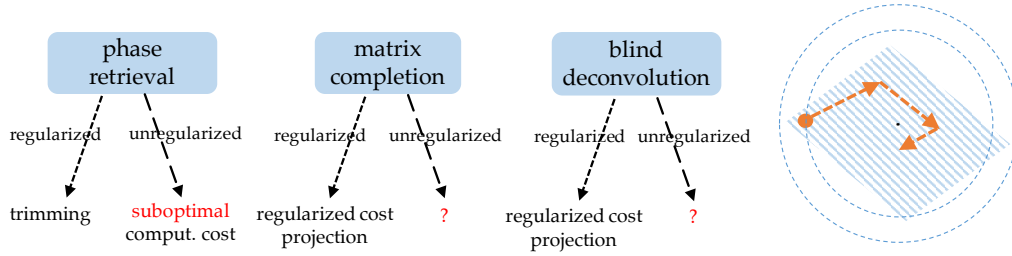


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.

AFOSR YIP Final Report
Taming Nonconvexity in Solving High-Dimensional Nonlinear
Systems of Equations
Award No.: FA9550-19-1-0030

PI: Yuxin Chen (Princeton University)

Research objectives. This research project aims to develop foundational theory and efficient algorithms that allow to solve high-dimensional nonlinear systems of equations in a time-critical, robust, and model-insensitive manner. This is a key enabler of efficient extraction of meaningful information and actionable intelligence from complex data sources in various modern data science applications. At the core of such challenges is the nonconvexity issue: a diverse array of such problems can naturally be formulated as nonconvex optimization problems. However, the dramatically increased dimensionality of data requires one to solve large-scale nonconvex optimization problems, which are computationally infeasible in general. Focusing on the design and analysis of problem-specific nonconvex optimization algorithms, an overall goal is to design a new suite of nonconvex optimization methods that possess both statistical and computational guarantees when accommodating the ever-increasing dimensionality of the problems.

Accomplishments. During the entire funding periods (i.e., 2019-2022) of this project, we continue to study nonconvex optimization algorithms for multiple fundamentally important nonconvex problems. Our investigation has led to the state-of-the-art results in multiple fundamentally important problems, including but not limited to solving random quadratic systems of equations, low-rank matrix completion, solving bilinear systems of equations, low-rank tensor completion, robust principal component analysis, and policy optimization in reinforcement learning. In what follows, several highlights of our accomplishments (including specific objectives and outcomes) are discussed in more detail.

- *Randomly initialized nonconvex algorithms when solving random quadratic equations.* An issue that often confuses researchers is the necessity 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 problem of solving random 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.

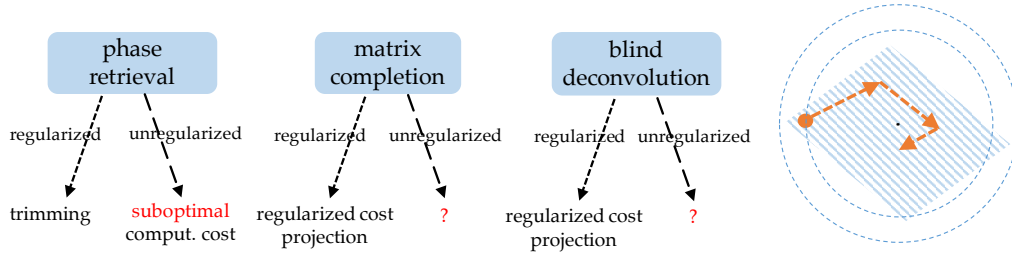


Figure 1: (Left) prior literature often requires sophisticated model-specific regularization in order to achieve fast convergence; (right) our results show that simple first-order methods implicitly regularize their iterates, thus achieving the fastest possible convergence without any explicit regularization.

The convergence guarantee we derive is at least $\text{poly}(n)$ 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.

- *Implicit regularization of nonconvex optimization algorithms.* In order to allow nonconvex optimization algorithms to be efficient for data science problems, 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 low-rank matrix completion, solving random quadratic systems of equations, solving bilinear systems of equations, and so on. In all of these problems, we discover that the algorithms implicitly regularize the running 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 accuracy.
- *Understanding the effectiveness of convex relaxation via nonconvex optimization in noisy matrix completion.* Apart from their computational advantages, nonconvex schemes also achieve optimal statistical accuracy under the presence of noise, which could be significantly better than those bounds derived for convex relaxation. Such a counter-intuitive theoretical discrepancy motivates us to take a closer inspection of the underlying connection between the two contrasting algorithmic frameworks. Focusing on the noisy matrix completion problem, we uncover an interesting and inherent connection: nonconvex optimization is essentially solving convex relaxation with nuclear norm penalization. Specifically, we discover that an approximate stationary point of the nonconvex formulation serves as an extremely tight approximation of the convex solution. This bridge allows us to transfer the optimal theoretical guarantees of nonconvex schemes to its convex counterpart, which yields the conclusion that

convex relaxation is also minimax optimal vis-a-vis random noise. This answers positively a conjecture posed by Candès and Plan in 2009.

- *Nonconvex low-rank tensor completion from noisy data.* Moving beyond the vector and matrix cases, we have investigated 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 — which is substantially more challenging than the matrix counterpart due to the notorious tensor structure. 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 (e.g., either requiring sub-optimal sample size or converging to a low-quality solution). 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 efficiency. 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 low-rank matrix factorization problems.* Modern decision making not merely anticipates a faithful estimate of the object under study, but also seeks to quantify the uncertainty or “confidence” of the provided estimate, ideally in a reasonably accurate fashion. This inferential reasoning provides scientists and practitioners with valuable and generalizable insights that can potentially lead to better decisions. Conducting effective uncertainty quantification for high-dimensional problems is, however, far from straightforward. For the most part, the state-of-the-art estimators require solving highly complex nonconvex optimization problems, which often do not admit closed-form solutions. Of necessity, it is generally very challenging to pin down the distributions of the estimates returned by these algorithms. The lack of distributional characterizations is a major roadblock to performing valid inference on the parameters of interest. In this project, we develop an effective uncertainty quantification toolkit for several popular nonconvex methods in high-dimensional problems, including noisy matrix completion and covariance estimation. In particular, we develop simple procedures to compensate for the bias of the widely used nonconvex estimators for both problems. The resulting de-biased estimators admit nearly precise non-asymptotic distributional characterizations, which in turn enable optimal construction of confidence intervals and confidence 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 characterization of the estimation accuracy of our de-biased estimators, which, to the best of our knowledge, are the first tractable algorithms that provably achieve full statistical efficiency (including the preconstant).
- *Fast global convergence of nonconvex optimization for reinforcement learning.* A remarkable 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 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.

- *Solving bilinear systems of equations via nonconvex optimization.* We investigate the effectiveness of nonconvex optimization in solving a system of bilinear equations (also referred to as blind deconvolution). This problem spans multiple domains in science and engineering, including but not limited to astronomy, medical imaging, optics, and communication engineering. Despite a large body of prior works tackling this problem, however, how to efficiently achieve optimal accuracy in the presence of random noise remains unsettled. Our recent work makes important contributions by demonstrating that: a carefully designed two-stage nonconvex optimization algorithm attains minimax-optimal accuracy within a logarithmic number of iterations. This convergence result accommodates both i.i.d. random designs and semi-random Fourier design; in particular, the latter one is large motivated by practical applications. Our results significantly improve upon the state-of-the-art theoretical guarantees. Our theoretical finding is largely based on an implicit regularization phenomenon that permeates nonconvex optimization algorithms, namely, many nonconvex algorithms automatically constrain the iterates to stay within some nice regions associated with benign optimization landscapes, thus enabling fast convergence.
- *Learning mixtures of low-rank models.* In this work, we consider the problem of learning mixtures of low-rank models, i.e., reconstructing a mixture of multiple low-rank matrices from unlabelled measurements of each. This problem enriches two widely studied settings — low-rank matrix sensing and mixed linear regression — by bringing latent variables (i.e., unknown labels) and structural priors (i.e., low-rank structures) into consideration. To cope with the non-convexity issues arising from unlabelled heterogeneous data and low-complexity structure, we develop a three-stage meta-algorithm that is guaranteed to recover the unknown matrices with near-optimal sample and computational complexities under Gaussian designs. In addition, the proposed algorithm is provably stable against random noise. We complement the theoretical studies with empirical evidence that confirms the efficacy of our algorithm.
- *Uncertainty quantification for nonconvex tensor completion.* 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.

- *Bridging convex and nonconvex optimization in robust PCA.* In this work, we consider the convex programming approach in low-rank matrix estimation, in the presence of (1) random noise, (2) gross sparse outliers, and (3) missing data. This problem, often dubbed as robust principal component analysis (robust PCA), finds applications in various domains. Despite the wide applicability of convex relaxation, the available statistical support (particularly the stability analysis vis-a-vis random noise) remains highly suboptimal, which we strengthen in this paper. For a broad class of ground-truth matrices, we demonstrate that a principled convex program achieves near-optimal statistical accuracy, in terms of both the Euclidean loss and the entrywise loss. All of this happens even when nearly a constant fraction of observations are corrupted by outliers with arbitrary magnitudes. All of this is enabled by bridging convex relaxation with the nonconvex Burer-Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise and outliers. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, thus allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

Dissemination. The PI has given more than 50 seminar talks in U.S. universities between 2019-2022 (e.g., Harvard university, Yale university, University of Pennsylvania, Princeton University, Stanford University, University of California at Berkeley, University of Southern California, Columbia University, Georgia Institute of Technology) to disseminate the research results obtained in this project. The PI has coauthored a monograph “Spectral methods for data science: a statistical perspective” covering various initialization schemes for nonconvex statistical estimation, published in the Foundations and Trends in Machine Learning series in 2021. In addition, the PI has also coauthored an invited overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which offers a contemporary overview of nonconvex optimization methods for low-rank statistical estimation problems.

Impacts.

- *Development of the principal discipline(s) of the project.* The research conducted in this project has led to the development of a new suite of algorithms that achieve optimal computational and statistical efficiency for a wide variety of nonconvex optimization problems.
- *Other disciplines.* The algorithms developed in this project are likely to be applicable and efficient for other applied disciplines beyond optimization, operations research, and statistics.
- *Describe the impact on the development of human resources.* Nothing to report.
- *Describe the impact on teaching and educational experiences.* The research results developed in this project have been incorporated into the PI’s two graduate courses: “large-scale optimization for data science,” and “mathematics of data science.” The course materials have been made publicly accessible. Based on these lecture materials, the PI has co-authored a monograph “Spectral methods for data science: a statistical perspective” and an overview article “Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview”, which have now been widely distributed. The PI has also cotaught a tutorial “Nonconvex Optimization for High-Dimensional Signal Estimation: Spectral and Iterative Methods” in EUSIPCO 2020.
- *Describe the impact on physical, institutional, and information resources that form infrastructure.* Nothing to report.
- *Impact on society beyond science and technology.* Nothing to report.

Changes. Nothing to report.

List of publications:

1. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 407-452, Jan. 2023.
2. “Nonconvex low-rank tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Operations Research*, vol. 70, no. 2, pp. 1219-1237, 2022.
3. Shicong Cen, Chen Cheng, Yuxin Chen, Yuting Wei, and Yuejie Chi, “Fast Global Convergence of Natural Policy Gradient Methods with Entropy Regularization,” *Operations Research*, vol. 70, no. 4, pp. 2563-2578, 2022.
4. Yuxin Chen, Jianqing Fan, Bingyan Wang, and Yuling Yan, “Convex and Nonconvex Optimization Are Both Minimax-Optimal for Noisy Blind Deconvolution,” accepted to *Journal of the American Statistical Association*.
5. “Learning Mixtures of Low-Rank Models,” Yanxi Chen, Cong Ma, H. Vincent Poor, and Yuxin Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 7, July 2021.
6. “Spectral Methods for Data Science: A Statistical Perspective,” Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, *Foundations and Trends in Machine Learning*, vol. 14, no. 5, pp. 566-806, 2021.
7. “Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data”, Y. Chen, Y. Chi, J. Fan, C. Ma, Y. Yan, *Annals of Statistics*, vol. 49, no. 5, pp. 2948-2971, Oct. 2021.
8. “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, Oct. 2020
9. “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
10. “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
11. “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**)
12. “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
13. “Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise”, C. Cheng, Y. Wei, Y. Chen, *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7380-7419, Nov. 2021.

14. “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
15. “Nonconvex low-rank symmetric tensor completion from noisy data”, C. Cai, G. Li, H. V. Poor, Y. Chen, *Neural Information Processing Systems*, 2019.
16. “Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality”, C. Cai, H. V. Poor, Y. Chen, *International Conference on Machine Learning*, 2020.
17. “Communication-efficient distributed optimization in networks with gradient tracking and variance reduction”, B. Li, S. Cen, Y. Chen, Y. Chi, *International Conference on Artificial Intelligence and Statistics*, 2020
18. “Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees”, C. Cai, G. Li, Y. Chi, H. V. Poor, Y. Chen, *the Annals of Statistics*, vol. 49, no. 2, pp. 944-967, 2021.
19. “Nonconvex Matrix Factorization from Rank-One Measurements,” Yuanxin Li, Cong Ma, Yuxin Chen, and Yuejie Chi, *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1928-1950, March 2021
20. “Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices”, Y. Chen, C. Cheng, J. Fan, *the Annals of Statistics*, vol. 49, no. 1, pp. 435-458, 2021
21. “Communication-Efficient Distributed Optimization in Networks with Gradient Tracking and Variance Reduction,” Boyue Li, Shicong Cen, Yuxin Chen, Yuejie Chi, *Journal of Machine Learning Research*, vol. 21, no. 180, pp. 1-51, 2020.