FAST QUADRATIC SENSING VIA NONCONVEX OPTIMIZATION

Kaihui Liu[†] *Liangtian Wan*[‡] *Lu Sun*[§]

[†] Great Bay University
[‡] School of Software, Dalian University of Technology
[§] Institute of Inf. Science Technology, Dalian Maritime University

ABSTRACT

This paper concerns the problem of recovering an unknown but rank-r signal from quadratic measurements. We formulate the recovery problem as two-stage nonconvex optimization approach where an efficient adaptive accelerated gradient-like iterative refinement starting from a careful initialization. Moreover, an efficient initial scheme is proposed for this problem. The average relative error of our initialization estimator reduces exponentially as the oversampling ratio grows, which can also boost the performance of existing two-stage approach for rank-r matrix recovery from vector measurements. Experimental results not only clearly demonstrate the superiority of introduced initialization estimator but also show the advantages of our adaptive accelerated gradient-like approach in terms of both sample complexity and computational complexity.

Index Terms— Adaptive accelerated gradient-like method, low-rank matrix recovery, nonconvex optimization, quadratic sensing.

1. INTRODUCTION

In a variety of applications including coherence retrieval in optical imaging [1] and quantum state tomography [2], covariance sketching of high-dimensional streaming data [3], one faces to recover a rank-r matrix $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ from a set of scalar quadratic measurements. This problem is also known as quadratic sensing [4] and the scalar measurements are given as

$$y_i = \|\boldsymbol{a}_i^H \boldsymbol{X}\|^2 + e_i,\ i = 1, ..., m \tag{1}$$

where $\{\boldsymbol{a}_i \in \mathbb{C}^n,\ i \le i \le m\}$ is a set of (known) sensing vectors, $e_i \in \mathbb{R}$ for $i = 1, ..., m$ represents the additive noise. Let the rank r matrix $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ be fixed and our interest is to recover $\boldsymbol{X}$ from a set of scalar quadratic measurements, i.e.,

$$\begin{aligned} &\text{find } \boldsymbol{X} \in \mathbb{C}^{n\times r} \\ &\text{s.t } y_i = \boldsymbol{a}_i^H \boldsymbol{X}\boldsymbol{X}^H \boldsymbol{a}_i + e_i,\ i = 1, ..., m \end{aligned} \tag{2}$$

The problem (2) is equivalent in recovering a rank-r positive semidefinite matrix $\boldsymbol{M} = \boldsymbol{X}\boldsymbol{X}^H \in \mathbb{C}^{n\times n}$ from m scalar quadratic measurements

$$\begin{aligned} &\text{find } \boldsymbol{M} \in \mathbb{C}^{n\times n} \\ &\text{s.t } y_i = \boldsymbol{a}_i^H \boldsymbol{M} \boldsymbol{a}_i + e_i,\ i = 1, ..., m \\ &\quad \text{rank}(\boldsymbol{M}) \le r. \end{aligned} \tag{3}$$

Among the most well-established methods based on semidefinite relaxation (e.g., [5], [6].) can solve the (3) with optimal sample complexity under Gaussian measurement. However, these methods often lift the original n-dimensional natural parameter space to a higher dimensional $n \times n$ space and cast the problem (3) as a semidefinite programming, which is, in general, computationally expensive to deal with large-scale data.

To avoid this drawback, many approaches focus on alternative formulations in the natural parameter space and solve the nonconvex optimization problem. In [7], the vanilla gradient descent, following a tailored spectral initialization, is proved for solving the rank-r minimization problem from a nonconvex quadratic loss function. A variational Bayesian learning approach is proposed for low rank phase retrieval [8] with unknown rank information. An exponential-type gradient descent algorithm is proposed in [9] to minimize the nonconvex quadratic loss function, which is similar to the truncation rule in [10] to suppress samples that heavily influence the search direction. But their method consider the nonconvex *intensity-based* quadratic loss function and the rank r matrix is assumed in real space, which is continuously differentiable. Recently, vanilla Gradient Descent is used for quadratic sensing under *amplitude-based* loss function [11]. Similar nonconvex method is proved to converge the ground truth (up to global ambiguity) with near-optimal sample complexity for rank-one measurements [12]. However, the conventional spectral initialization can't provide a high quality

This work was supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515111193, National Natural Science Foundation of China (62101088), Radar Signal Processing National Defense Science and Technology Key Laboratory Fund (6142401200101) and Dalian High-level Talent Innovation Support Plan No. 2019RQ024.
Corresponding author: Kaihui Liu (email: KaihuiL@outlook.com).

initial guess [13, 14], which leads to all these nonconvex optimization methods using conventional spectral initialization may not show their best performance.

To overcome these limitations, we utilize the amplitude-based loss function inspired by the importance of better models in nonconvex phase retrieval and further develop an efficient and powerful nonconvex optimization-based method for the quadratic sensing problem.

2. QUADRATIC SENSING PROBLEM

Following the least-square criterion, the problem of (2) can be naturally recast as the ensuing empirical loss minimization, and a nonconvex optimization based iterative method can be derived in the natural parameter space:

$$\boldsymbol{X} \leftarrow \arg \min_{\boldsymbol{U} \in \mathbb{C}^{n \times r}} L(\boldsymbol{U}) \triangleq \frac{1}{m} \sum_{i=1}^{m} \ell(\boldsymbol{U}; y_i) \tag{4}$$

where the loss function $L(\boldsymbol{U})$ is selected based on the data measurements and the noise model. Similar to [15], [16], [8], the *amplitude-based* loss function is utilized for this paper as follows:

$$L(\boldsymbol{U}) \triangleq \frac{1}{m} \sum_{i=1}^{m} \left(\sqrt{y_i} - \|\boldsymbol{a}_i^H \boldsymbol{U}\|_2\right)^2, \forall\, i \in \{1, ..., m\} \tag{5}$$

However, the objective function $L(\boldsymbol{U})$ is nonconvex. Minimizing nonconvex objectives is NP-hard in general due to existence of many stationary points [17]. Fortunately, the empirical loss often enjoys benign geometry in a small local region surrounding the global optimum and specially, starting with a carefully designed initial guess that is in a basin of attraction of $\boldsymbol{U}$, successive refinement is effected based upon a sequence of gradient-like iterations without leaving the basin. This two-stage approach often leads to very efficient algorithms that run in time proportional given data dimension [4]. Select a learning rate μ, start with an initial point $\boldsymbol{U}^{(0)} \in \mathbb{C}^{n \times r}$, and generate a sequence $\{\boldsymbol{U}^{(k+1)} \subseteq \mathbb{C}^{n \times r}\}$ via the iteration

$$\boldsymbol{U}^{(k+1)} = \boldsymbol{U}^{(k)} - \frac{\mu^{(k)}}{m} \sum_{i \in \mathcal{H}^{(k+1)}} \nabla \ell_i(\boldsymbol{U}^{(k)}; y_i) \tag{6}$$

where $\boldsymbol{U}^{(k)}$ denotes the estimate returned by the algorithm at the k-th iteration, $\mu^{(k)} > 0$ is the step size, $\nabla \ell_i(\boldsymbol{U}^{(k)}; y_i)$ is the (generalized) gradient-like of loss function $L(\boldsymbol{U})$, and $\mathcal{H}^{(k+1)}$ represents certain selected index set effecting the per-iteration. If $L(\boldsymbol{U})$ is a mapping from $\mathbb{C}^n$ to $\mathbb{R}$, it is not holomorphic and hence not complex-differentiable. However, it is viewed as a gradient based on Wirtinger derivatives in [18], and the Wirtinger derivative of $L(\boldsymbol{U}; y_i, \varepsilon)$ is computed as

$$\nabla L(\boldsymbol{U}; y_i) := \frac{2}{m} \sum_{i=1}^{m} (\|\boldsymbol{a}_i^H \boldsymbol{U}\|_2 - \sqrt{y_i}) \times \frac{\boldsymbol{a}_i \boldsymbol{a}_i^H \boldsymbol{U}}{\|\boldsymbol{a}_i^H \boldsymbol{U}\|_2} \tag{7}$$

(7) is also similar to the (generalized) gradient-like method derived from [19] as

$$\nabla \ell_i(\boldsymbol{U}^{(k)}; y_i) = \left(\|\boldsymbol{a}_i^H \boldsymbol{U}^{(k)}\|_2 - \sqrt{y_i}\right) \frac{\boldsymbol{a}_i \boldsymbol{a}_i^H \boldsymbol{U}}{\|\boldsymbol{a}_i^H \boldsymbol{U}\|_2} \tag{8}$$

3. INITIALIZATION

A key to the success of general nonconvex optimization method is an effective initialization. The quality of the start point for the nonconvex optimization method may significantly affect the performance of iterative searching process. If one can find a high quality start point that is in a local basin of attraction of ground truth $\boldsymbol{X}$, an iterative refinement procedure would find the wanted solution of (4). This two-stage approach is remarkably efficient under Gaussian measurements starting with a carefully-designed initialization estimators and step-size rules. In [7], a tailored spectral initialization is proposed for providing the initial guess for nonconvex optimization, which exploit the principal subspace of a surrogate matrix constructed by the measurements and sensing vectors to provide an initialization. An argument similar to phase retrieval can be applied to quadratic sensing in (4), recognizing that the expectation of the surrogate matrix as

$$\boldsymbol{D} = \frac{1}{2m} \sum_{i=1}^{m} y_i \boldsymbol{a}_i \boldsymbol{a}_i^H \tag{9}$$

and under the Gaussian design, the surrogate matrix $\boldsymbol{D}$ can be viewed as the sample average of m i.i.d. random rank-one matrices $\{y_i \boldsymbol{a}_i \boldsymbol{a}_i^H\}_{i=1}^m$ [12]. When the number of samples m is large, this sample average should be "close"to its expectation. The spectral method then proceeds by computing $\boldsymbol{Z} \in \mathbb{C}^{n \times r}$ (which consists of the top-r eigenvectors of $\boldsymbol{D}$), and a diagonal matrix $\boldsymbol{\Lambda}$ whose i-th diagonal value is given as

$$\boldsymbol{\Lambda}_{ii} = \lambda_i - \lambda, \;\; i = 1, ..r \tag{10}$$

where $\lambda = \lambda_{r+1}$ [7] (But in [12], λ is suggested as $\frac{1}{m}\sum_{i=1}^{m} y_i$ provides a more accurate estimate.). The initial guess is then set as

$$\boldsymbol{U}_0 = \boldsymbol{Z} \boldsymbol{\Lambda}^{1/2} \tag{11}$$

3.1. Modified Spectral Initialization

In [9], a truncation rule similar to [10] is introduced to spectral initialization for rank-r quadratic sensing problem, where a high-quality initial guess scheme is provided and just need $O(nr)$ measurements. Motivated by the success of pre-processing function [13, 14, 20] used in phase retrieval. We explored similar tricks to the quadratic sensing problem and the pre-processing function $\mathcal{T}(\cdot)$ used in this paper derived by [21]. It should be noted that the other excellent pre-processing functions [20, 22] also work efficiently.

$$\mathcal{T}(y_i) \triangleq \left(c_1 - \exp\left(-y_i^p / \xi^2\right)\right), \; i \in I \tag{12}$$

where $p \in (0, 1]$ (we choose $p = 1$ through the paper) and the constant c_1 is defined as

$$c_1 \leftarrow \begin{cases} \frac{\sqrt{3}}{3} & \text{if } \boldsymbol{X} \in \mathbb{R}^{n\times r}, \\ \frac{1}{2} & \text{if } \boldsymbol{X} \in \mathbb{C}^{n\times r}. \end{cases} \tag{13}$$

and the normalized constant is

$$\xi \leftarrow \sqrt{\frac{1}{|I|}\sum_{i\in I} y_i} \tag{14}$$

The positive value of pre-processing function $\mathcal{T}(y_i)$, $i \in \{1, m\}$ is used to extract useful information from the large values of $\{y_i\}_{i=1}^m$. Small measurements y_i, $i \in \{1, m\}$ is regarded as "weak" signal [23] in which the sensing vectors $\boldsymbol{a}_i$, $i \in \{1, m\}$ is nearly orthogonal with the targeted ground truth $\boldsymbol{X}$. Evidently, pre-processing of the "weak "signal $\mathcal{T}(y_i)$ may have a large negative part which is can be considered as penalizing the "weak" measurements and hence preserves the information of targeted ground truth $\boldsymbol{X}$. Similar to the spectral initialization (9), the new surrogate matrix is defined as

$$\tilde{\boldsymbol{D}} \triangleq \frac{1}{2m}\sum_{i=1}^{m} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^H \tag{15}$$

and under the Gaussian design, the expectation of surrogate matrix $\tilde{\boldsymbol{D}}$ can be viewed as,

$$\mathbb{E}[\tilde{\boldsymbol{D}}] = 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H + \|\tilde{\boldsymbol{X}}\|_F^2\boldsymbol{I}_n \tag{16}$$

The spectral method then proceeds by computing $\tilde{\boldsymbol{Z}} \in \mathbb{C}^{n\times r}$, which are the normalized eigenvectors corresponding to the r largest eigenvalues of $\tilde{\boldsymbol{D}}$, and a diagonal matrix $\tilde{\boldsymbol{\Lambda}}$ whose i-th diagonal value is given as

$$\tilde{\boldsymbol{\Lambda}}_{ii} = \tilde{\lambda}_i - \tilde{\lambda}, \;\; i = 1, ..r \tag{17}$$

where $\tilde{\lambda} = \frac{1}{m}\sum_{i=1}^m y_i$. Therefore, the new initial guess is given by

$$\boldsymbol{U}_0 = \tilde{\boldsymbol{Z}}\tilde{\boldsymbol{\Lambda}}^{1/2} \tag{18}$$

4. OPTIMIZATION ALGORITHMS

Although the gradient-like iterative algorithm is computational efficiency, the convergence rate of gradient-like method is $\mathcal{O}(1/k)$, which means that $\mathcal{O}(k)$ iterations are required to reach a wanted $\mathcal{O}(1/k)$ accuracy of solution. Recently, Nesterov's accelerated scheme is proved that can escape saddle points and has faster convergence rate than gradient descent in the nonconvex setting [24]. Inspired by [25], [26], we introduce a fast adaptive restart gradient-like approach starting with our proposed initialization to update (6).

The $(k + 1)$-th iteration of fast adaptive restart gradient-like method is updated as

$$\boldsymbol{U}^{(k+1)} = \boldsymbol{U}^{(k)} - \frac{\mu^{(k)}}{m}\sum_{i\in\{m\}} \nabla \ell_i(\boldsymbol{U}^{(k)}; y_i) \tag{19}$$

$$\begin{aligned}\boldsymbol{V}^{k+1} =& \boldsymbol{U}^{k+1} + \frac{\eta^k - 1}{\eta^{k+1}}(\boldsymbol{U}^{k+1} - \boldsymbol{U}^k) \\ &+ \frac{\eta^k}{\eta^{k+1}}(\boldsymbol{U}^{k+1} - \boldsymbol{V}^k)\end{aligned} \tag{20}$$

with

$$\eta^{k+1} = \begin{cases} \frac{\sqrt{4(\eta^k)^2+1}+1}{2}, & k \le T-1 \\ \frac{\sqrt{8(\eta^k)^2+1}+1}{2}, & k = T \end{cases} \tag{21}$$

and η^1 is set to be 1.

Unlike gradient-like method, accelerated schemes are not guaranteed to be monotone in the loss function. Restarting the accelerated algorithm from the current iteration is a common technique is applied in accelerated approaches. We would suggest an elegant approach named adaptive restart technique [26] to implement Algorithm 1. The restarting operation is activated whenever

$$L(\boldsymbol{U}^{(k+1)}) > L(\boldsymbol{U}^{(k)}) \tag{22}$$

For clarity, we summarize our algorithm as Algorithm 1.

Algorithm 1 Adaptive Accelerated Gradient Descent Method (AAGD)

Input: Sensing vectors $\{\boldsymbol{a}_i\}_{i=1}^m$, data $\boldsymbol{y}$, learning rate μ, the maximum number of iterations T.

1. **Initialization:** $\boldsymbol{U}_0$ is obtained from modified spectral initialization.
2. Let $\boldsymbol{V}^1 = \boldsymbol{U}^1 = \boldsymbol{U}_0, \eta^1 = 1, k = 1$.
3. **while** (not meet the stop criterion) **do**
   (a) Update $\boldsymbol{U}^{k+1}$ according to (19).
   (b) Update $\boldsymbol{V}^{k+1}$ from (20) and η^{k+1} from (21).
4. **end while**

Output: $\hat{\boldsymbol{X}} = \boldsymbol{U}^{T+1}$.

5. NUMERICAL SIMULATIONS

In this section, we conduct a number of experiments to evaluate the performance of introduced modified spectral initialization (MSI) and a fast adaptive restart gradient-like approach (AAGD) starting with our proposed MSI initialization. Our proposed methods are compared with existing state-of-the-art (nonconvex) algorithms for the quadratic sensing problem.

For fairness, all procedures are implemented with their suggested parameter values. We take a Gaussian sensing vectors $\boldsymbol{a}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_n) \triangleq \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n/2) + j\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n/2), 1 \le i \le m$ and generate a rank r matrix $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ from an independent standard multivariate normal distribution and where $\boldsymbol{X}_{i,i} \sim \mathcal{CN}(0,1), 1 \le i \le n, 1 \le j \le r$. All the experiments are implemented in Matlab 2016b and carried out on a computer equipped with Intel Core i5 3.4GHz CPU and 8GB RAM. All simulated results reported in this paper were averaged over 100 Monte Carlo (MC) realizations. The performance evaluation metric used relative error defined as:

$$\text{RE} \leftarrow \min_{\boldsymbol{O}\in\mathcal{O}(r)} \frac{\|\boldsymbol{X}\boldsymbol{O} - \hat{\boldsymbol{X}}\|_F}{\|\boldsymbol{X}\|_F} = \frac{\|\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{V}^H - \hat{\boldsymbol{X}}\|_F}{\|\boldsymbol{X}\|_F} \quad (23)$$

where $\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{V}^H$ is the singular value decomposition of $\boldsymbol{X}^H\hat{\boldsymbol{X}}$, $\hat{\boldsymbol{X}}$ is the estimate of targeted rank-r matrix. Fig. 1 shows

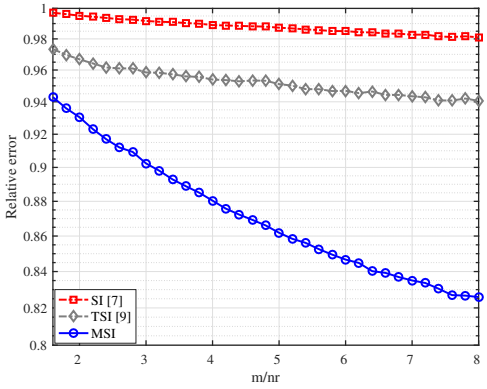

Fig. 1: Average relative error for $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ with $n = 50, r = 4$ under the complex Gaussian model and m/n varying within the range $[1.6, 8]$.

that the average *relative error* of the modified spectral initialization (MSI) is compared with truncated spectral initialization (TSI) [9], spectral initialization (SI) [7], [12], [11] for $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ with $n = 50, r = 4$ under the complex Gaussian model and m/nr varying within the range $[1.6, 8]$. We observe that our introduced modified spectral initialization yields a lower average relative error and significantly outperforms the existing spectral initialization methods as the oversampling ratio m/nr grows.

Fig. 2 depicts the empirical success rate of our AAGD method, compared with GD method [7], [12], [11], and AltMin method [11] versus varying oversampling ratio $m/nr \in [1.8, 3.2]$ for $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ with $n = 50$ and $r = 4$ under the complex Gaussian model, where a success trial is claimed if the estimate has a relative error less than 10^{-5}. Note that it was shown in [9] that the exponential-type GD method empirically on a par with the GD method [7] and assume the real Gaussian model. Therefore, the Exponential-type GD method was not included for comparison. For boosting the performance of compared method, all the methods start with the

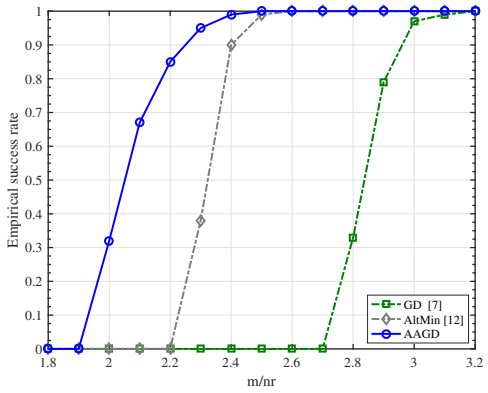

Fig. 2: Empirical success rate versus oversampling ratio m/nr for $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ with $n = 50, r = 4$.

proposed MSI initialization. Our proposed AAGD method can significantly improve the performance of Vanilla Gradient Descent used in [7], [12], [11].

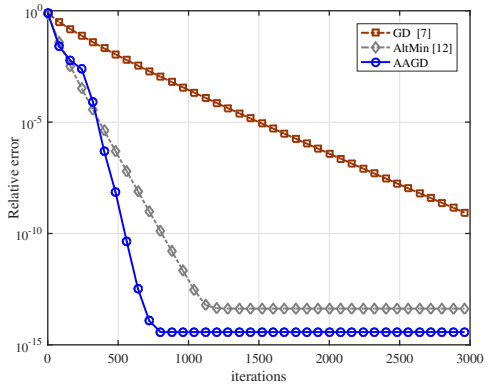

Fig. 3: The NMSE with respect to iterations for $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ with $n = 50, r = 4$ and $m = 6n$.

Fig. 3 further compares the convergence speed of various schemes in terms of the number of iterations to produce solutions of a given accuracy. Evidently, AAGD method converge faster than vanilla gradient descent method used in [7], [12], [11] and AltMin method [11], and all methods test under the complex Gaussian model with $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ with $n = 50$, $r = 4$ and $m = 6n$.

6. CONCLUSION

In this paper, we propose an efficient initialization estimator and an adaptive accelerated gradient-like approach starting with proposed initialization for solving quadratic sensing problem. The average relative error of our initialization estimator reduces exponentially as the oversampling ratio grows, which can boost the performance of existing two-stage approach for rank-r matrix recovery from vector measurements.

Experimental results not only clearly demonstrate the superiority of introduced initialization estimator but also show the advantages of our adaptive accelerated gradient-like approach in terms of both sample complexity and computational complexity.

7. REFERENCES

[1] C. Bao, G. Barbastathis, H. Ji, Z. Shen, and Z. Zhang, "Coherence retrieval using trace regularization," *SIAM J. Imag. Sci.*, vol. 11, no. 1, pp. 679–706, 2018.

[2] R. Kueng, H. Rauhut, and U. Terstiege, "Low rank matrix recovery from rank one measurements," *Appl. Comput. Harmon. Anal.*, vol. 42, no. 1, pp. 88–116, 2017.

[3] Y. Chen, Y. Chi, and A. J. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4034–4059, 2015.

[4] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5239–5269, 2019.

[5] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[6] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Programm.*, vol. 149, no. 1-2, pp. 47–81, 2015.

[7] S. Sanghavi, R. Ward, and C. D. White, "The local convexity of solving systems of quadratic equations," *Results Math.*, vol. 71, no. 3-4, pp. 569–608, Jun. 2017.

[8] K. Liu, J. Wang, Z. Xing, L. Yang, and J. Fang, "Low-rank phase retrieval via variational bayesian learning," *IEEE Access*, vol. 7, pp. 5642–5648, 2019.

[9] M. Huang and Z. Xu, "Solving systems of quadratic equations via exponential-type gradient descent algorithm," *J. Comput. Math.*, vol. 38, no. 4, pp. 638–660, 2020.

[10] Y. Chen and E. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Comm. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–883, 2017.

[11] V. Monardo, Y. Li, and Y. Chi, "Solving quadratic equations via amplitude-based nonconvex optimization," in *Proc. Int. Conf. Acoust. Speech. Signal Process.*, 2019, pp. 5526–5530.

[12] Y. Li, C. Ma, Y. Chen, and Y. Chi, "Nonconvex matrix factorization from rank-one measurements," *IEEE Trans. Inf. Theory*, vol. 67, no. 3, pp. 1928–1950, 2021.

[13] K. Liu, L. Li, and L. Wan, "Weighted null vector initialization and its application to phase retrieval," in *Proc. Int. Conf. Acoust. Speech. Signal Process.*, 2020, pp. 4861–4865.

[14] K. Liu, L. Yang, and J. Xie, "Robust initialization estimator and its application in solving quadratic equations," *IEEE Access*, vol. 7, 2019.

[15] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms," *J. Mach. Learn. Res.*, vol. 18, no. 141, pp. 1–35, 2017.

[16] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 773–794, 2018.

[17] A. Ben-Tal, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Society for Industrial and Applied Mathematics :, 2001.

[18] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[19] F. H. Clarke, "Generalized gradients and applications," *Trans. Amer. Math. Soc.*, vol. 205, p. 247–262, Apr. 1975.

[20] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *Found. Comput. Math.*, vol. 19, no. 3, pp. 703–773, 2019.

[21] B. Gao and Z. Xu, "Phaseless recovery using the gauss-newton method," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5885–5896, Nov. 2017.

[22] Y. M. Lu and G. Li, "Phase transitions of spectral initialization for high-dimensional non-convex estimation," *Inf. Inference, J. IMA*, vol. 9, no. 3, pp. 507–541, 2020.

[23] P. Chen, A. Fannjiang, and G. R. Liu, "Phase retrieval by linear algebra," *SIAM J. Matrix Anal. Appl.*, vol. 38, no. 3, pp. 854–868, 2017.

[24] C. Jin, P. Netrapalli, and M. I. Jordan, "Accelerated gradient descent escapes saddle points faster than gradient descent," in *Proc. Conf. Comput. Learn. Theory*, Stockholm, Sweden, 2018, pp. 1–44.

[25] D. Kim and J. A. Fessler, "Optimized first-order methods for smooth convex minimization," *Math. Programm.*, vol. 159, no. 1-2, pp. 81–107, 2016.

[26] B. O'donoghue and E. Candes, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, vol. 15, no. 3, pp. 715–732, 2015.