SOLVING QUADRATIC EQUATIONS VIA AMPLITUDE-BASED NONCONVEX OPTIMIZATION

Vincent Monardo, Yuanxin Li and Yuejie Chi

Department of Electrical and Computer Engineering
Carnegie Mellon University
Emails: {vmonardo,yuanxinl,yuejiec}@andrew.cmu.edu

ABSTRACT

In many signal processing tasks, one seeks to recover an r-column matrix object $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ from a set of nonnegative quadratic measurements up to orthonormal transforms. Example applications include coherence retrieval in optical imaging and covariance sketching for high-dimensional streaming data. To this end, efficient nonconvex optimization methods are quite appealing, due to their computational efficiency and scalability to large-scale problems. There is a recent surge of activities in designing nonconvex methods for the special case $r = 1$, known as phase retrieval; however, very little work has studied the general rank-r setting. Motivated by the success of phase retrieval, in this paper we derive several algorithms which utilize the quadratic loss function based on amplitude measurements, including (stochastic) gradient descent and alternating minimization. Numerical experiments demonstrate their computational and statistical performances, highlighting the superior performance of stochastic gradient descent with appropriate mini-batch sizes.

***Index Terms*—** Low-rank matrix recovery, quadratic measurement, nonconvex optimization

1. INTRODUCTION

In this paper, we are interested in recovering an r-column matrix object $\boldsymbol{X} \in \mathbb{C}^{n\times r}$ from a set of nonnegative quadratic measurements, given as

$$y_i = \|\boldsymbol{a}_i^H \boldsymbol{X}\|_2^2, \quad i = 1, \ldots, m, \tag{1}$$

where $\boldsymbol{a}_i \in \mathbb{C}^n$ is the ith sensing vector, and m is the number of measurements. In the Gaussian design, the sensing vectors are generated i.i.d. from a complex-valued Gaussian distribution, i.e. $\boldsymbol{a}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$. Equivalently, this problem amounts to recovering a rank-r positive semidefinite matrix $\boldsymbol{M} = \boldsymbol{X}\boldsymbol{X}^H \in \mathbb{C}^{n\times n}$ from a set of linear measurements

$$y_i = \boldsymbol{a}_i^H \boldsymbol{M} \boldsymbol{a}_i = \langle \boldsymbol{M}, \boldsymbol{a}_i \boldsymbol{a}_i^H \rangle, \tag{2}$$

where the sensing matrix $\boldsymbol{a}_i \boldsymbol{a}_i^H$ is rank-one. Since the measurements $\boldsymbol{y} = \{y_i\}_{i=1}^m$ are nonnegative, they are dubbed *phaseless* measurements. The goal is to recover $\boldsymbol{M}$, or equivalently the factor $\boldsymbol{X}$ up to orthonormal transforms from as few measurements $m \ll n^2$ as possible. This problem arises in many applications, ranging from coherence retrieval in optical imaging [1], covariance sketching of high-dimensional streaming data [2, 3] for the general rank-r case, to phase retrieval [4–7] for the rank-1 case. The measurements $\boldsymbol{y}$ are quadratic in both $\boldsymbol{a}_i$'s and $\boldsymbol{X}$, and we therefore also refer to this sensing model as *quadratic sensing*.

There are two lines of approaches to solve this problem. The first one is based on convex relaxations to solve for the low-rank matrix $\boldsymbol{M}$ [5, 8]. These algorithms perfectly recover the underlying matrix $\boldsymbol{M}$ at a near-optimal sample complexity under the Gaussian design. However, the computational complexity of the resulting semidefinite programs scale at least cubically with the dimension of $\boldsymbol{M}$, and therefore is prohibitive when the problem size is large.

This leads to the second line of approaches, which are iterative algorithms based on nonconvex optimization that directly estimate the factor $\boldsymbol{X}$ [9–12]. In [13, 14], it is proposed to recover $\boldsymbol{X}$ directly by minimizing the following loss function that is the squared error of the *intensity* measurements y_i:

$$\ell_{\text{in}}(\boldsymbol{U}) := \frac{1}{m}\sum_{i=1}^m \left(y_i - \|\boldsymbol{a}_i^H \boldsymbol{U}\|_2^2\right)^2, \tag{3}$$

where $\boldsymbol{U} \in \mathbb{C}^{n\times r}$. The algorithm proposed in [13, 14] is gradient descent with spectral initialization, and its efficiency is proved in [14] for the Gaussian design.

1.1. Our contributions

This paper proposes to solve for $\boldsymbol{X}$ by minimizing the squared error of the *amplitude* measurements, that is, $z_i = \sqrt{y_i}$ for $i = 1, ..., m$. The goal is to minimize the following loss function:

$$\ell(\boldsymbol{U}) = \frac{1}{m}\sum_{i=1}^m \left(z_i - \|\boldsymbol{a}_i^H \boldsymbol{U}\|_2\right)^2, \tag{4}$$

which is both nonconvex and nonsmooth. To the best of our knowledge, this loss function has not been considered to solve the general rank-r quadratic sensing problem considered here. Compared with the intensity-based loss function (3), the amplitude-based loss function (4) is a lower-order polynomial with respect to $\boldsymbol{a}_i^H \boldsymbol{U}$, and therefore is expected to have better curvatures around the global optimum, and more amenable to fast computation. We developed three algorithms to optimize (4): gradient descent (GD), mini-batch stochastic gradient descent (SGD), and alternating minimization, which are initialized by the spectral method based on amplitude measurements. All of these algorithms converge to a critical point of $\ell(\boldsymbol{U})$, and empirically achieve stronger statistical and computational performances than optimizing the intensity-based loss function (3) via gradient descent [14].

This work is supported in part by ONR under the grant N00014-18-1-2142, by ARO under the grant W911NF-18-1-0303, and by NSF under grants CCF-1806154 and ECCS-1818571.

1.2. Related Work

This work is motivated by the successful adoption of the amplitude-based loss function in phase retrieval [15–17], where it achieves near-optimal computational and statistical complexities without requiring sophisticated truncation or regularization procedures as using the intensity-based loss function [10]. Besides [13, 14], an exponential-type gradient descent algorithm is proposed in [18] to minimize (3), which is similar to the truncation rule in [10] to suppress samples that heavily influence the search direction. A few papers proposed other algorithms to solve the quadratic sensing problem, including but not limited to [19, 20], but they are applied to either the lifted formulation (2) or the intensity-based based loss function (3).

1.3. Paper Organization and Notations

The rest of this paper is organized as follows. Section 2 presents the proposed algorithms using the amplitude-based loss function, including (stochastic) gradient descent and alternating minimization. Section 3 examines and provides numerical comparisons of the proposed algorithms with existing approaches. Finally, we conclude in Section 4.

We use boldfaced symbols to represent vectors and matrices. For any vector $\boldsymbol{v}$, we let $\|\boldsymbol{v}\|_2$ denote the ℓ_2 norm. For any matrix $\boldsymbol{M}$, we let $\|\boldsymbol{M}\|_F$ denote the Frobenius norm. In addition, we use $\boldsymbol{M}^H$ and $\boldsymbol{M}^\dagger$ to indicate the conjugate transpose and the pseudo-inverse of $\boldsymbol{M}$, respectively. The diagonal matrix with the diagonal entries given by the vector $\boldsymbol{v}$ is denoted as $\text{diag}(\boldsymbol{v})$.

2. AMPLITUDE-BASED NONCONVEX OPTIMIZATION

We start by providing the intuition behind using the amplitude-based loss function (4) and then introduce three different algorithms to minimize it: gradient descent, stochastic gradient descent which also utilizes mini-batches, and alternating minimization. Finally, we propose a spectral initialization based on the amplitude measurements.

2.1. Making Sense of the Amplitude-Based Loss Function

We start by defining a generalized "phase" vector

$$\boldsymbol{b}_i^H = \frac{\boldsymbol{a}_i^H \boldsymbol{X}}{\|\boldsymbol{a}_i^H \boldsymbol{X}\|_2} \in \mathbb{C}^r \tag{5}$$

corresponding to each phaseless measurement. Then with the amplitude measurements $z_i = \|\boldsymbol{a}_i^H \boldsymbol{X}\|_2$, we can write a set of linear measurements of $\boldsymbol{X}$ as

$$\boldsymbol{a}_i^H \boldsymbol{X} = \boldsymbol{b}_i^H z_i, \qquad i = 1, ..., m.$$

Furthermore, define

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^H \\ \boldsymbol{a}_2^H \\ \vdots \\ \boldsymbol{a}_m^H \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1^H \\ \boldsymbol{b}_2^H \\ \vdots \\ \boldsymbol{b}_m^H \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}. \tag{6}$$

Then we can compactly write

$$\boldsymbol{A}\boldsymbol{X} = \text{diag}(\boldsymbol{z})\boldsymbol{B}. \tag{7}$$

Indeed if the matrix $\boldsymbol{B}$ is known, then $\boldsymbol{X}$ can be solved via standard least-squares. The challenge is that the phase term $\boldsymbol{B}$ is unknown and we cannot apply least-squares directly. To this end, we aim to find the phase matrix $\boldsymbol{B}$ and $\boldsymbol{X}$ that minimize the loss function:

$$\min_{\boldsymbol{U}, \|\boldsymbol{p}_i\|_2 = 1} \ell(\boldsymbol{U}, \boldsymbol{P}) = \frac{1}{m} \|\boldsymbol{A}\boldsymbol{U} - \text{diag}(\boldsymbol{z})\boldsymbol{P}\|_F^2, \tag{8}$$

under the constraint that the rows of $\boldsymbol{P}$ are unit-norm. Interestingly, when fixing $\boldsymbol{U}$ (the estimate of $\boldsymbol{X}$), the phase vector that minimizes the right-hand side of (8) can be found in a closed-form as

$$\boldsymbol{p}_i^H = \boldsymbol{a}_i^H \boldsymbol{U} / \|\boldsymbol{a}_i^H \boldsymbol{U}\|_2. \tag{9}$$

Plugging (9) into $\ell(\boldsymbol{U}, \boldsymbol{P})$ lead to the amplitude-based loss function in (4):

$$\ell(\boldsymbol{U}) = \min_{\|\boldsymbol{p}_i\|_2 = 1} \ell(\boldsymbol{U}, \boldsymbol{P}).$$

In words, the amplitude-based loss function can be regarded as an attempt to approximate the least-squares loss in the absence of the phase information.

Algorithm 1 Gradient Descent with Amplitude Loss

Input: $\{z_i\}_{i=1}^m$, $\{\boldsymbol{a}_i\}_{i=1}^m$, step size μ_k
Initialization:
Obtain $\boldsymbol{U}_0$ from Spectral Initialization (Algorithm 4)
Gradient Updates:
for $k = 0, 1, 2, ..., K-1$ **do**
$\boldsymbol{W}_k = \text{diag}\left(\left[\frac{\|\boldsymbol{a}_1^H \boldsymbol{U}_k\|_2 - z_1}{\|\boldsymbol{a}_1^H \boldsymbol{U}_k\|_2}, ..., \frac{\|\boldsymbol{a}_m^H \boldsymbol{U}_k\|_2 - z_m}{\|\boldsymbol{a}_m^H \boldsymbol{U}_k\|_2}\right]\right)$
$\boldsymbol{U}_{k+1} = \boldsymbol{U}_k - \frac{2\mu_k}{m} \boldsymbol{A}^H \boldsymbol{W}_k \boldsymbol{A} \boldsymbol{U}_k,$
end for
return $\boldsymbol{U}_K$

2.2. Gradient Descent

A first approach is to apply gradient descent to minimize (4), which may proceed at each iteration $k \geq 0$ as

$$\boldsymbol{U}_{k+1} = \boldsymbol{U}_k - \mu_k \nabla \ell(\boldsymbol{U}_k) \tag{10}$$

for some step size μ_k, and $\boldsymbol{U}_0$ is a properly chosen initialization that will be discussed later. Due to the nonsmoothness of the loss function, the generalized gradient [21] of $\ell(\boldsymbol{U})$ with respect to $\boldsymbol{U}$ is used:

$$\nabla \ell(\boldsymbol{U}) = \frac{2}{m} \sum_{i=1}^m \left(\|\boldsymbol{a}_i^H \boldsymbol{U}\|_2 - z_i\right) \frac{\boldsymbol{a}_i \boldsymbol{a}_i^H \boldsymbol{U}}{\|\boldsymbol{a}_i^H \boldsymbol{U}\|_2}$$

The details of gradient descent with the amplitude loss function is given in Algorithm 1. Comparing with the least-squares case when the generalized phase matrix $\boldsymbol{B}$ is known, here at each iteration, the phase vector is estimated via the current iterate.

2.3. Mini-Batch Stochastic Gradient Descent

Next, we implement a stochastic version of gradient descent using mini-batches, which is found in practice to be compelling both statistically and computationally. By utilizing a stochastic method with an appropriate mini-batch size, we significantly decrease the computation cost per iteration while still converging at a moderate number of iterations. The details of the mini-batch stochastic gradient descent (SGD) is given in Algorithm 2.

Algorithm 2 Mini-Batch SGD with Amplitude Loss

Input: $\{z_i\}_{i=1}^m$, $\{\boldsymbol{a}_i\}_{i=1}^m$, mini-batch size B, step size μ_k
Initialization:
Obtain $\boldsymbol{U}_0$ from Spectral Initialization (Algorithm 4)
Gradient Updates:
for $k = 0, 1, 2, ..., K-1$ **do**
Choose Γ_k uniformly at random from $\{1, 2, ..., m\}$ with cardinality B and update
$\boldsymbol{W}_{\Gamma_k} = \text{diag}\left(\left\{ \frac{\|\boldsymbol{a}_i^H \boldsymbol{U}_k\|_2 - z_i}{\|\boldsymbol{a}_i^H \boldsymbol{U}_k\|_2} \,\middle|\, i \in \Gamma_k \right\}\right)$
$\boldsymbol{U}_{k+1} = \boldsymbol{U}_k - \frac{2\mu_k}{m} \boldsymbol{A}_{\Gamma_k}^H \boldsymbol{W}_{\Gamma_k} \boldsymbol{A}_{\Gamma_k} \boldsymbol{U}_k$,
where $\boldsymbol{A}_{\Gamma_k}$ is a matrix stacking of $\boldsymbol{a}_i^H$ for $i \in \Gamma_k$ as its rows.
end for
return $\boldsymbol{U}_K$

Algorithm 3 Alternating Minimization

Input: $\{z_i\}_{i=1}^m$, and $\{\boldsymbol{a}_i\}_{i=1}^m$
Initialization:
Obtain $\boldsymbol{U}_0$ from Spectral Initialization (Algorithm 4)
Alternating Updates:
for $k = 0, 1, 2, ..., K-1$ **do**
$\boldsymbol{U}_{k+1} = \boldsymbol{A}^\dagger \text{diag}\left(\left[\frac{z_1}{\|\boldsymbol{a}_1^H \boldsymbol{U}_k\|_2}, ..., \frac{z_m}{\|\boldsymbol{a}_m^H \boldsymbol{U}_k\|_2}\right]\right) \boldsymbol{A}\boldsymbol{U}_k$
end for
return $\boldsymbol{U}_K$

2.4. Alternating Minimization

Last but not least, we propose alternating minimization (AltMin) to update the phase matrix and the signal sequentially to minimize $\ell(\boldsymbol{U}, \boldsymbol{P})$ in (8), which leads to a direct generalization of the well-known Gerchberg-Saxton algorithm for phase retrieval [22] to the general quadratic sensing problem. In words, at each iteration $k \geq 0$, we update each row of the phase matrix $\boldsymbol{P}_k = [\boldsymbol{p}_1^k, \ldots, \boldsymbol{p}_m^k]^H$ as

$$(\boldsymbol{p}_i^k)^H = \boldsymbol{a}_i^H \boldsymbol{U}_k / \|\boldsymbol{a}_i^H \boldsymbol{U}_k\|_2, \quad i = 1, \ldots, m, \tag{11}$$

Then, we update $\boldsymbol{U}_{k+1}$ by fixing $\boldsymbol{P}_k$ and solving a least-squares problem:

$$\begin{aligned} \boldsymbol{U}_{k+1} &= \text{argmin}_{\boldsymbol{U} \in \mathbb{C}^{n \times r}} \|\boldsymbol{A}\boldsymbol{U} - \text{diag}(\boldsymbol{z})\boldsymbol{P}_k\|_2^2 \\ &= \boldsymbol{A}^\dagger \,\text{diag}(\boldsymbol{z})\boldsymbol{P}_k \\ &= \boldsymbol{A}^\dagger \,\text{diag}\left(\left[\frac{z_1}{\|\boldsymbol{a}_1^H \boldsymbol{U}_k\|_2}, ..., \frac{z_m}{\|\boldsymbol{a}_m^H \boldsymbol{U}_k\|_2}\right]\right) \boldsymbol{A}\boldsymbol{U}_k. \end{aligned} \tag{12}$$

The details of the alternating minimization algorithm is shown in Algorithm 3. It is easy to check that this update rule is guaranteed to not increase the amplitude loss function at every iteration:

$$\begin{aligned} \ell(\boldsymbol{U}_{k+1}) &= \min_{\|\boldsymbol{p}_i\|=1} \frac{1}{m} \sum_{i=1}^m \|\boldsymbol{a}_i^H \boldsymbol{U}_{k+1} - y_i \boldsymbol{p}_i\|_2^2 \\ &\leq \frac{1}{m} \|\boldsymbol{A}\boldsymbol{U}_{k+1} - \text{diag}(\boldsymbol{y})\boldsymbol{P}_k\|_2^2 \\ &\leq \frac{1}{m} \|\boldsymbol{A}\boldsymbol{U}_k - \text{diag}(\boldsymbol{y})\boldsymbol{P}_k\|_2^2 \\ &= \ell(\boldsymbol{U}_k), \end{aligned}$$

where the second inequality follows from (12).

2.5. Spectral Initialization

So far, all of the algorithms require an initialization $\boldsymbol{U}_0$, which hopefully is close to the ground truth $\boldsymbol{X}$ that we wish to recover. The spectral method is a popular method to provide a high-quality initial guess in nonconvex optimization, where we construct a data matrix based on the measurements and sensing vectors and use its principal subspace to provide an initial guess. In this paper, we advocate the use of amplitude measurements to construct the data matrix, as detailed in Algorithm 4. The method is inspired by [14], except for the use of amplitude measurements. Numerical experiments in the later section will verify the advantage of this approach.

Algorithm 4 Spectral Initialization with Amplitude Measurements

Input: $\{z_i\}_{i=1}^m$, and $\{\boldsymbol{a}_i\}_{i=1}^m$
Define the data matrix $\boldsymbol{D} = \frac{1}{2m} \sum_{i=1}^m z_i \boldsymbol{a}_i \boldsymbol{a}_i^H$.
Obtain the r normalized eigenvectors $\boldsymbol{Z}_0 \in \mathbb{C}^{n \times r}$ corresponding to the r largest eigenvalues of $\boldsymbol{D}$.
Obtain the diagonal matrix $\boldsymbol{\Lambda}_0 \in \mathbb{C}^{r \times r}$, with entries on the diagonal given by

$$[\boldsymbol{\Lambda}_0]_i = \lambda_i(\boldsymbol{D}) - \lambda, \quad i = 1, ..., r$$

where $\lambda = \frac{1}{m} \sum_{i=1}^m z_i$ and $\lambda_i(\boldsymbol{D})$ is the i^{th} largest eigenvalue of $\boldsymbol{D}$.
return $\boldsymbol{U}_0 = \boldsymbol{Z}_0 \boldsymbol{\Lambda}_0^{1/2}$.

3. NUMERICAL EXPERIMENTS

In this section, we provide the empirical performance of various amplitude-based algorithms that we consider, with comparisons to gradient descent based on the intensity-based loss function [14]. We will first compare their statistical performance in terms of sample complexities, and then compare their computational performance in terms of wall-clock time taken to achieve a desired accuracy. Extensive experiments are conducted over a wide range of settings of problem dimensions; here we report the most representative results that we find are consistent over different problem instances.

For each run, the entries of the sensing vectors $\boldsymbol{a}_i$ are generated i.i.d. using complex-valued Gaussian variables $\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$. The performance is measured using the Normalized Mean Squared Error (NMSE), defined as

$$\text{NMSE} = \frac{\|\boldsymbol{U}\boldsymbol{U}^H - \boldsymbol{X}\boldsymbol{X}^H\|_F}{\|\boldsymbol{X}\boldsymbol{X}^H\|_F}, \tag{13}$$

where $\boldsymbol{U}$ is the estimate of an algorithm. For GD and mini-batch SGD with the amplitude loss function, we use a constant step size of $\mu = 0.8$, whereas for GD with the intensity loss function, we use a constant step size of $\mu = 0.13$, which are optimized for convergence consistency. Additionally, we pick a mini-batch size of $B = \lceil \frac{n}{3} \rceil$ for SGD, which performs the best in terms of convergence time.

3.1. Comparisons of Spectral Initialization Methods

We start by comparing the performance of spectral initializations using amplitude and intensity measurements. In order to justify the used of amplitude measurements with spectral initialization, we ran an experiment of $10,000$ trials to compare the empirical distribution of NMSE using the amplitude-based and intensity-based spectral initialization, as shown in Fig. 1. The amplitude-based spectral initialization provides a lower NMSE and has a much lower variance than

the intensity-based spectral initialization. For the rest of the simulations, we will use the amplitude-based spectral initialization.

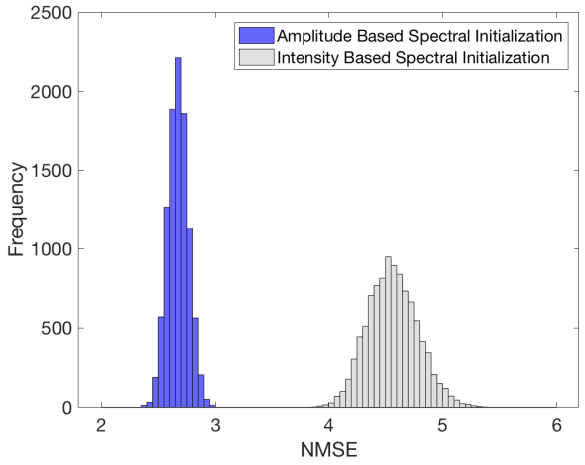

Fig. 1. The empirical distribution of NMSE for spectral initialization constructed using amplitude and intensity measurements.

3.2. Comparisons of Statistical Performance

An important metric for the proposed algorithms is the minimum number of measurements required in order to recover the ground truth. Fix $n = 50$ and $r = 4$. We vary the number of measurements and measure the empirical success rate of each algorithm over 50 Monte Carlo simulations. A trial is labeled as a success when NMSE $\leq 10^{-5}$. Fig. 2 shows the empirical success rate with respect to the sampling ratio $m/(nr)$ using the same spectral initialization, for GD using the amplitude-based loss, GD using the intensity-based loss, mini-batch SGD using the amplitude-based loss, and alternating minimization. All the algorithms achieve perfect recovery as long as the sampling ratio is large enough; moreover, all of three proposed algorithms experience a phase transition using fewer measurements than GD using the intensity-based loss, indicating the benefit of the amplitude-based loss. Finally, the mini-batch SGD outperforms the rest of the algorithms, which is consistent with the observation in the phase retrieval case [15].

3.3. Comparisons of Computational Performance

We next compare the computational performance of each algorithm, by demonstrating the decrease in NMSE as time passes. We track the convergence in terms of wall-clock time as opposed to the number of iterations to avoid the confounding variable of time-per-iteration. All of experiments were run on a MacBook Air with a 2.2 GHz Intel Core i7 and 8 GB of 1600 MHz DDR3 RAM.

Fig. 3 shows the NMSE with respect to the wall-clock time when $n = 50$, $r = 4$, and $m = 800$ using the same spectral initialization for the same set of algorithms as in Fig. 2. Again, to achieve the same accuracy in terms of NMSE, GD using the intensity-based loss requires more time than all three algorithms using the amplitude-based loss. In particular, the mini-batch SGD requires significantly less time than GD or alternating minimization, due to its much faster execution per iteration.

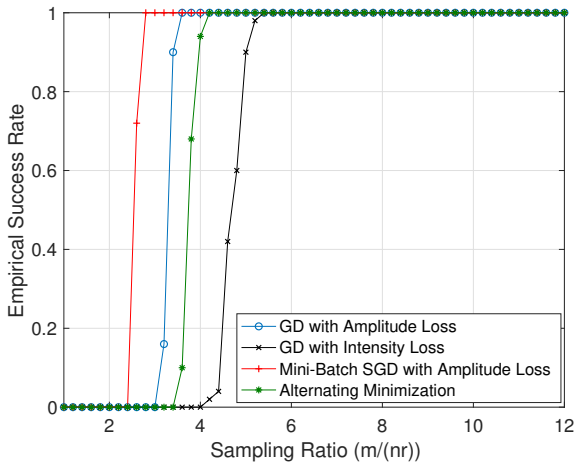

Fig. 2. The empirical success rate with respect to the sampling ratio $m/(nr)$ for various algorithms, when $n = 50$ and $r = 4$.

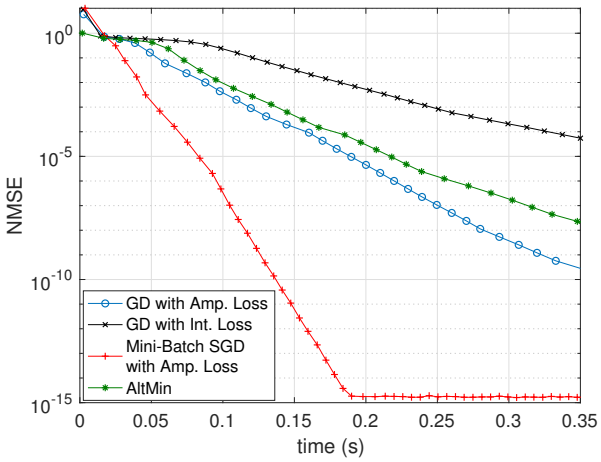

Fig. 3. The NMSE with respect to wall-clock time for various algorithms, when $n = 50$, $r = 4$ and $m = 800$.

4. CONCLUSIONS

In this paper, we propose three new nonconvex approaches for solving quadratic sensing, or equivalently, low-rank matrix recovery from rank-one measurements, which all utilize an amplitude-based loss function as opposed to an intensity-based loss function. Numerical experiments are provided to demonstrate their advantages over using the intensity-based loss function in terms of both sample complexity and computational complexity.

In the future, we would like to provide theoretical analysis to guarantee the convergence of each of these algorithms to the global optimum solution as long as the sample size is sufficiently large. Another interesting venue of research would be to demonstrate the stability of these algorithms to noise and/or the presence of outliers in the measurements [23] using random initialization [24].

5. REFERENCES

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

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

[3] T. T. Cai and A. Zhang, "Rop: Matrix recovery via rank-one projections," *The Annals of Statistics*, vol. 43, no. 1, pp. 102–138, 2015.

[4] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

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

[6] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

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

[8] Y. Li, Y. Sun, and Y. Chi, "Low-rank positive semidefinite matrix recovery from corrupted rank-one measurements," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 397–408.

[9] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[10] Y. Chen and E. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Advances in Neural Information Processing Systems (NIPS)*, 2015.

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

[12] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *arXiv preprint arXiv:1809.09573*, 2018.

[13] S. Sanghavi, R. Ward, and C. D. White, "The local convexity of solving systems of quadratic equations," *Results in Mathematics*, pp. 1–40, 2016.

[14] Y. Li, C. Ma, Y. Chen, and Y. Chi, "Nonconvex matrix factorization from rank-one measurements," *arXiv preprint arXiv:1802.06286*, 2018.

[15] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5164–5198, 2017.

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

[17] J. Ma, J. Xu, and A. Maleki, "Approximate message passing for amplitude based optimization," *arXiv preprint arXiv:1806.03276*, 2018.

[18] M. Huang and Z. Xu, "Solving systems of quadratic equations via exponential-type gradient descent algorithm," *arXiv preprint arXiv:1806.00904*, 2018.

[19] K. Zhong, P. Jain, and I. S. Dhillon, "Efficient matrix sensing using rank-1 Gaussian measurements," in *International Conference on Algorithmic Learning Theory*. Springer, 2015, pp. 3–18.

[20] M. Soltani and C. Hegde, "Improved algorithms for matrix recovery from rank-one projections," *arXiv preprint arXiv:1705.07469*, 2017.

[21] A. Y. Kruger, "On Fréchet subdifferentials," *Journal of Mathematical Sciences*, vol. 116, no. 3, pp. 3325–3358, 2003.

[22] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237, 1972.

[23] H. Zhang, Y. Chi, and Y. Liang, "Provable non-convex phase retrieval with outliers: Median truncated Wirtinger flow," in *International conference on machine learning*, 2016, pp. 1022–1031.

[24] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval," *arXiv preprint arXiv:1803.07726*, 2018.