Kaczmarz Method for Solving Quadratic Equations

Yuejie Chi, *Member, IEEE* and Yue M. Lu, *Senior Member, IEEE*

***Abstract*—Estimating low-rank positive-semidefinite matrices from symmetric rank-one measurements is of great importance in many applications, such as high-dimensional data processing, quantum state tomography, and phase retrieval. When the rank is known a priori, this problem can be regarded as solving a system of quadratic equations of a low-dimensional subspace. We develop a fast iterative algorithm based on an adaptation of the Kaczmarz method, which is traditionally used for solving over-determined linear systems. In particular, we characterize the dynamics of the algorithm when the measurement vectors are composed of standard Gaussian entries in the online setting. Numerical simulations demonstrate the compelling performance of the proposed algorithm.**

***Index Terms*—Quadratic equations, Kaczmarz method, low-rank matrix recovery, online algorithms**

I. Introduction

Recent years have witnessed a surge of research activities in provably recovering a low-rank matrix from a number of generic linear measurements much smaller than its ambient dimension, using both convex and non-convex procedures [1]–[4]. In this paper, we are interested in estimating a low-rank positive semidefinite (PSD) matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ from a number of its linear measurements in the form:

$$z_i = \langle \mathbf{A}_i, \mathbf{\Sigma}\rangle = \langle \mathbf{a}_i\mathbf{a}_i^T, \mathbf{\Sigma}\rangle = \mathbf{a}_i^T\mathbf{\Sigma}\mathbf{a}_i, \quad i = 1,\ldots,m, \quad (1)$$

where the measurement matrix $\mathbf{A}_i = \mathbf{a}_i\mathbf{a}_i^T \in \mathbb{R}^{n\times n}$ is PSD and rank-one with $\mathbf{a}_i \in \mathbb{R}^n$, and m is the number of measurements with $m \ll n^2$. This model is motivated by many applications such as quantum space tomography [5], covariance estimation [6], phase retrieval [7], and projection retrieval [8]. Moreover, compared with generic cases when $\mathbf{A}_i$ is full-rank, the computational and storage complexity of the measurement process (1) is much lower.

To date, several works have considered the reconstruction of $\mathbf{\Sigma}$ from (1) using convex relaxation techniques, by solving either a trace norm minimization algorithm [6], [9] or a feasibility problem [10] under both the PSD constraint and the measurement constraint. In this work, it is established that the matrix $\mathbf{\Sigma}$ can be exactly recovered as soon as the number of measurements m exceeds the order of nr when $\mathbf{a}_i$'s are composed of i.i.d. standard Gaussian entries, where r is the rank of $\mathbf{\Sigma}$. This sample complexity is near-optimal information theoretically. However, these convex algorithms have a running time that is cubic in n [11], which becomes prohibitive when n is large.

Motivated by the fact that if the rank (or its upper bound) of $\mathbf{\Sigma}$ is known a priori, it is possible to represent $\mathbf{\Sigma}$ by its Cholesky decomposition, i.e. $\mathbf{\Sigma} = \mathbf{U}\mathbf{U}^T$, where $\mathbf{U} \in \mathbb{R}^{n\times r}$ is of much smaller dimension than $\mathbf{\Sigma}$. We reformulate (1) as a set of *quadratic equations* in $\mathbf{U}$, given as

$$z_i = \mathbf{a}_i^T\mathbf{U}\mathbf{U}^T\mathbf{a}_i = \left\|\mathbf{U}^T\mathbf{a}_i\right\|_2^2, \quad i = 1,\ldots,m. \quad (2)$$

The Wirtinger Flow (WF) algorithm [12], [13] is proven as an effective non-convex algorithm to recover $\mathbf{U}$ from (2) for phase retrieval when $r = 1$, which is later generalized to the general low-rank setting in [14]. The WF algorithm starts with a careful initialization using the spectral method [11], and then updates the estimate via gradient descent using either least-squares or Poisson loss functions. Global convergence properties of WF are established as soon as the number of measurements m is on the order of n when $r = 1$ [13], and on the order of $nr^\alpha \log^2 n$ for some small positive integer $\alpha \geq 1$ for the general low-rank case [14]. The computational complexity of WF is linear in both m and n, making it computationally much more attractive than the convex approaches.

The goal of this paper is to propose a new fast iterative algorithm that directly estimates $\mathbf{U}$ from the quadratic system (2), based on an adaptation of the well-known Kaczmarz method [15], originally developed as an efficient method for solving over-determined *linear* systems. Recently, the Kaczmarz method has received renewed attention due to its marriage to randomization tricks [16]–[18] that yields provable global convergence properties. In [19], [20], the Kaczmarz method is applied to phase retrieval by integrating a phase selection heuristic. Extensive empirical comparisons are conducted in [19] to demonstrate the computational advantage of the Kaczmarz method over state-of-the-art methods such as WF, and its large-system limit is characterized in [20] when $\mathbf{a}_i$'s are composed of standard Gaussian entries.

We systematically derive the Kaczmarz method to solve the general low-rank case in (2) by mimicking the update rule in the linear case: in each iteration, we seek the closest matrix to the current estimate satisfying the sampled measurement constraint. Interestingly, for our quadratic model in (2), this can be solved efficiently with a simple closed-form solution by resorting to the Woodbury matrix inversion formula. Moreover, the update rule in each iteration corresponds to a rank-one update of the subspace $\mathbf{U}$, whose computational complexity is linear in n. Numerical simulations are provided to validate the effectiveness of the proposed algorithm, where it

Y. Chi is with Department of Electrical and Computer Engineering and Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio 43210. Email: chi.97@osu.edu. The work of Y. Chi is supported in part by grants NSF CCF-1422966 and AFOSR FA9550-15-1-0205.

Y. M. Lu is with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. Email: yuelu@seas.harvard.edu. The work of Y. M. Lu was supported in part by NSF under grant CCF-1319140 and by ARO under grant W911NF-16-1-0265. Part of this work was done during his stay at the Information Initiative at Duke (iiD). He thanks this interdisciplinary program for hospitality and support.

compares favorably over the WF algorithm. The dynamics of the Kaczmarz algorithm is briefly analyzed for the Gaussian measurement case, showing that the dynamics can be fully captured by a random walk with $2r^2$ parameters, irrespective of the underlying ambient dimension n.

The rest of the paper is organized as follows. Section II describes the proposed Kaczmarz algorithm. Section III provides some preliminary analysis of the dynamics of the Kaczmarz algorithm in the online setting. Numerical simulations are presented in Section IV to demonstrate its superior performance. Finally, we conclude in Section V.

II. The Proposed Algorithm

Define $y_i = z_i^{1/2}$, where z_i is given in (2), as

$$y_i = z_i^{1/2} = \left\|\mathbf{U}^T\mathbf{a}_i\right\|_2, \quad i = 1, \ldots, m, \tag{3}$$

where $\mathbf{U} \in \mathbb{R}^{n\times r}$. For each measurement, only the norm of $\mathbf{U}^T\mathbf{a}_i$ is observed without its orientation. When $r = 1$, the problem is equivalent to phase retrieval. Our goal is to recover $\mathbf{U}$ from $\{y_i\}_{i=1}^m$.

A. Kaczmarz method for linear systems

Before describing the proposed algorithm, recall the original Kaczmarz method for solving over-determined linear system:

$$h_i = \mathbf{a}_i^T\mathbf{x}, \quad i = 1, \ldots, m.$$

At the kth iteration, first pick an index $1 \le \ell(k) \le m$ (either randomly or sequentially), and then update the estimate of $\mathbf{x}$ as the vector satisfying the sampled measurement constraint that is closest to the previous estimate $\mathbf{x}_{k-1}$:

$$\begin{aligned}\mathbf{x}_k &= \operatorname{argmin}_{\mathbf{w}:h_{\ell(k)}=\mathbf{a}_{\ell(k)}^T\mathbf{w}} \left\|\mathbf{x}_{k-1} - \mathbf{w}\right\|_2^2 \\ &= \mathbf{x}_{k-1} + \left(\frac{h_{\ell(k)} - \mathbf{a}_{\ell(k)}^T\mathbf{x}_{k-1}}{\|\mathbf{a}_{\ell(k)}\|_2^2}\right)\mathbf{a}_{\ell(k)}.\end{aligned}$$

The Kaczmarz method can be implemented very efficiently by only accessing one measurement constraint per iteration.

B. Kaczmarz method for quadratic systems in (3)

Motivated by the updating rule in the linear case, at each iteration, we similarly first pick an index $1 \le \ell(k) \le m$ (either uniformly at random or sequentially), and update the next estimate as the solution to the following problem:

$$\mathbf{U}_k = \operatorname{argmin}_{\mathbf{X}:\left\|\mathbf{X}^T\mathbf{a}_{\ell(k)}\right\|_2=y_{\ell(k)}} \left\|\mathbf{U}_{k-1} - \mathbf{X}\right\|_{\mathrm{F}}^2, \tag{4}$$

which seeks the matrix satisfying the quadratic constraint that is closest to the previous estimate in the Frobenius norm. Fortunately, the above problem admits a simple closed-form solution that facilitates its implementation, as given in the following proposition.

Proposition 1. *The solution to* (4) *is given as*

$$\mathbf{U}_k = \left[\mathbf{I} - \left(\frac{\|\mathbf{U}_{k-1}^T\mathbf{a}_{\ell(k)}\|_2 - y_{\ell(k)}}{\|\mathbf{U}_{k-1}^T\mathbf{a}_{\ell(k)}\|_2}\right)\frac{\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T}{\|\mathbf{a}_{\ell(k)}\|_2^2}\right]\mathbf{U}_{k-1}. \tag{5}$$

Proof. Denote the Lagrangian of (4) as

$$\mathcal{L}(\mathbf{X}, \lambda) = \|\mathbf{U}_{k-1} - \mathbf{X}\|_{\mathrm{F}}^2 + \lambda\left(\mathbf{a}_{\ell(k)}^T\mathbf{X}\mathbf{X}^T\mathbf{a}_{\ell(k)} - y_{\ell(k)}^2\right),$$

where λ is Lagrange multiplier associated with the equality constraint. Setting the derivatives of $\mathcal{L}(\mathbf{X}, \lambda)$ with respect to $\mathbf{X}$ and λ to zero, we have

$$\frac{\partial\mathcal{L}(\mathbf{X},\lambda)}{\partial\mathbf{X}} = 2\left(\mathbf{X} - \mathbf{U}_{k-1}\right) + 2\lambda\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T\mathbf{X} = 0, \tag{6}$$

$$\frac{\partial\mathcal{L}(\mathbf{X},\lambda)}{\partial\lambda} = \mathbf{a}_{\ell(k)}^T\mathbf{X}\mathbf{X}^T\mathbf{a}_{\ell(k)} - y_{\ell(k)}^2 = 0. \tag{7}$$

From (6), we can solve for $\mathbf{X}$ and obtain

$$(\mathbf{I} + \lambda\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T)\mathbf{X} = \mathbf{U}_{k-1},$$

which gives

$$\mathbf{X} = (\mathbf{I} + \lambda\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T)^{-1}\mathbf{U}_{k-1}. \tag{8}$$

Plugging (8) into (7) we obtain:

$$\begin{aligned}\mathbf{a}_{\ell(k)}^T(\mathbf{I} + \lambda\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T)^{-1}\mathbf{U}_{k-1}\mathbf{U}_{k-1}^T(\mathbf{I} + \lambda\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T)^{-1}\mathbf{a}_{\ell(k)} \\ = y_{\ell(k)}^2.\end{aligned} \tag{9}$$

Applying the Woodbury's matrix inversion formula

$$(\mathbf{I}+\lambda\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T)^{-1} = \mathbf{I} - \left(\frac{1}{\lambda} + \|\mathbf{a}_{\ell(k)}\|_2^2\right)^{-1}\mathbf{a}_{\ell(k)}\mathbf{a}_{\ell(k)}^T \tag{10}$$

to (9), we can solve for λ which gives

$$\lambda = \left(\pm\frac{\|\mathbf{U}_{k-1}^T\mathbf{a}_{\ell(k)}\|_2}{y_{\ell(k)}} - 1\right)\frac{1}{\|\mathbf{a}_{\ell(k)}\|_2^2}.$$

Note that because we need the solution that minimizes $\|\mathbf{X} - \mathbf{U}_{k-1}\|_{\mathrm{F}}$, we take the solution corresponding to the larger λ, yielding the update rule in (5). □

Remark 1. An alternative proof of Prop. 1 can be obtained by considering the following optimization for the "generalized" unknown phase $\mathbf{s} \in \mathbb{R}^r$, $\|\mathbf{s}\|_2 = 1$ of the measurement:

$$\begin{aligned}&\min_{\|\mathbf{s}\|_2=1} \min_{\mathbf{X}^T\mathbf{a}_{\ell(k)}=y_{\ell(k)}\mathbf{s}} \|\mathbf{U}_{k-1} - \mathbf{X}\|_{\mathrm{F}}^2 \\ &= \min_{\|\mathbf{s}\|_2=1} \frac{1}{\|\mathbf{a}_{\ell(k)}\|_2^2}\left\|y_{\ell(k)}\mathbf{s} - \mathbf{U}_{k-1}^T\mathbf{a}_{\ell(k)}\right\|_2^2,\end{aligned}$$

where the phase is found as $\mathbf{s} = \mathbf{U}_{k-1}^T\mathbf{a}_{\ell(k)}/\|\mathbf{U}_{k-1}^T\mathbf{a}_{\ell(k)}\|_2$ using the Cauchy-Schwarz inequality.

From (5), in each iteration we apply a rank-one update to the previous estimate, which admits a low computational complexity of $O(nr)$. Moreover, there is no tuning parameter. Algorithm 1 summarizes the proposed algorithm, where $\mathbf{U}_0$ is initialized either randomly or using the spectral method in [14]. Let $\mathbf{S} = \frac{1}{2m}\sum_{i=1}^m y_i^2\mathbf{a}_i\mathbf{a}_i^T$, whose top-$r$ eigenvectors and eigenvalues are given as $\mathbf{V} \in \mathbb{R}^{n\times r}$ and $\mathbf{\Lambda} = \operatorname{diag}(\{\lambda_t\}_{t=1}^r)$, we then set $\mathbf{U}_0 = \mathbf{V}\mathbf{\Lambda}^{1/2}$. Applying the spectral initialization does require access to all the data.

In particular, for the special phase retrieval case $r = 1$, the update rule (5) reduces to

$$\mathbf{u}_k = \mathbf{u}_{k-1} - \left[\frac{\mathbf{u}_{k-1}^T\mathbf{a}_{\ell(k)} - \operatorname{sign}(\mathbf{u}_{k-1}^T\mathbf{a}_{\ell(k)})y_{\ell(k)}}{\|\mathbf{a}_{\ell(k)}\|_2^2}\right]\mathbf{a}_{\ell(k)},$$

which is exactly the same as the one proposed in [19], [20].

Algorithm 1 Kaczmarz method for solving (3)

Input: $\{y_i\}_{i=1}^m$, $\{\mathbf{a}_i\}_{i=1}^m$, the rank r, and the number of iterations T.

Initialization: initialize $\mathbf{U}_0 \in \mathbb{R}^{n\times r}$ either randomly or using the spectral method.

1: **for** $k = 1, 2, \cdots, T$ **do**
2: Select an index $1 \le \ell(k) \le m$ uniformly at random;
3: Update $\mathbf{U}_k$ using (5);
4: **end for**

Output: $\hat{\mathbf{U}} = \mathbf{U}_T$.

III. Dynamic Analysis in the Online Setting

In the online case, we assume in each iteration a new measurement arrives independently, where $\ell(k) = k$. The update rule can be simplified as

$$\mathbf{U}_k = \mathbf{U}_{k-1} - \left(\frac{\|\mathbf{U}_{k-1}^T\mathbf{a}_k\|_2 - y_k}{\|\mathbf{U}_{k-1}^T\mathbf{a}_k\|_2}\right)\frac{\mathbf{a}_k\mathbf{a}_k^T\mathbf{U}_{k-1}}{\|\mathbf{a}_k\|_2^2}. \quad (11)$$

We will show that the dynamics of the proposed algorithm can be determined by a Markov process with a much smaller number of parameters than the ambient dimension $\mathbb{R}^{nr}$. Define the estimation error

$$d(\mathbf{U}_k, \mathbf{U}) = \operatorname{argmin}_{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}} \|\mathbf{U}_k - \mathbf{U}\mathbf{Q}\|_{\mathrm{F}}^2,$$

which can be computed as

$$\begin{aligned} d(\mathbf{U}_k, \mathbf{U}) &= \|\mathbf{U}_k\|_{\mathrm{F}}^2 + \|\mathbf{U}\|_{\mathrm{F}}^2 - 2\operatorname{argmax}_{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}} \operatorname{Tr}(\mathbf{U}_k^T\mathbf{U}\mathbf{Q}) \\ &= \|\mathbf{U}_k\|_{\mathrm{F}}^2 + \|\mathbf{U}\|_{\mathrm{F}}^2 - 2\|\mathbf{U}_k^T\mathbf{U}\|_1, \end{aligned}$$

where $\|\mathbf{U}_k^T\mathbf{U}\|_1 = \sum_{i=1}^r \lambda_i(\mathbf{U}_k^T\mathbf{U})$ is the sum of the singular values of $\mathbf{U}_k^T\mathbf{U}$, i.e. the nuclear norm. The last equality follows from the Von Neumann's trace inequality [21]. This suggests the performance is fully determined by the following two quantities:

$$b_k = \|\mathbf{U}_k\|_{\mathrm{F}}^2, \quad c_k = \|\mathbf{U}_k^T\mathbf{U}\|_1.$$

Define $\mathbf{\Phi}_k = \mathbf{U}_k^T\mathbf{U}_k \in \mathbb{R}^{r\times r}$, and $\mathbf{\Psi}_k = \mathbf{U}_k^T\mathbf{U} \in \mathbb{R}^{r\times r}$, then $b_k = \operatorname{Tr}(\mathbf{\Phi}_k)$ and $c_k = \|\mathbf{\Psi}_k\|_1$. From (11), it is possible to recursively update $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$. For $\mathbf{\Phi}_k$,

$$\begin{aligned} \mathbf{\Phi}_k &= \mathbf{\Phi}_{k-1} - \left(\frac{\|\mathbf{U}_{k-1}^T\mathbf{a}_k\|_2^2 - y_k^2}{\|\mathbf{U}_{k-1}^T\mathbf{a}_k\|_2^2}\right)\frac{\mathbf{U}_{k-1}^T\mathbf{a}_k\mathbf{a}_k^T\mathbf{U}_{k-1}}{\|\mathbf{a}_k\|_2^2} \\ &= \mathbf{\Phi}_{k-1} - \left(\frac{\|\mathbf{w}_k\|_2^2 - \|\mathbf{v}_k\|^2}{\|\mathbf{w}_k\|_2^2}\right)\frac{\mathbf{w}_k\mathbf{w}_k^T}{\|\mathbf{a}_k\|_2^2}, \end{aligned} \quad (12)$$

where $\mathbf{w}_k = \mathbf{U}_{k-1}^T\mathbf{a}_k$ and $\mathbf{v}_k = \mathbf{U}^T\mathbf{a}_k$. Similarly, for $\mathbf{\Psi}_k$,

$$\mathbf{\Psi}_k = \mathbf{\Psi}_{k-1} - \left(\frac{\|\mathbf{w}_k\|_2 - \|\mathbf{v}_k\|_2}{\|\mathbf{w}_k\|_2}\right)\frac{\mathbf{w}_k\mathbf{v}_k^T}{\|\mathbf{a}_k\|_2^2}. \quad (13)$$

Note that conditioned on $\mathbf{\Phi}_{k-1}, \mathbf{\Psi}_{k-1}$, $\mathbf{w}_k, \mathbf{v}_k$ are jointly Gaussian with the following distribution:

$$\mathbf{w}_k, \mathbf{v}_k | \mathbf{\Phi}_{k-1}, \mathbf{\Psi}_{k-1} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{\Phi}_{k-1} & \mathbf{\Psi}_{k-1} \\ \mathbf{\Psi}_{k-1}^T & \mathbf{U}^T\mathbf{U} \end{bmatrix}\right),$$

which in turn determines $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$. Therefore, $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$ form a Markov process with $2r^2$ parameters. We have the following proposition.

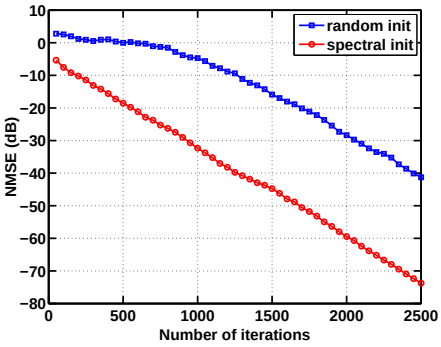


Fig. 1. The NMSE with respect to the number of iterations with a single pass of all measurements using the spectral initialization or the random initialization, when $n = 50$, $m = 2500$ and $r = 3$.

Proposition 2. *$\{\mathbf{\Phi}_k, \mathbf{\Psi}_k\}$ forms a Markov process in $\mathbb{R}^{2r^2}$.*

Due to space limit, we omit the formal proof of this proposition, which essentially follow similar arguments given in [20, Proposition 1]. This proposition provides a dimension-reduced representation of the error evolution of the proposed Kaczmarz method in the online setting, irrespective of the ambient dimension of $\mathbf{U}$. From (12) and (13), it is possible to find the large-system limit by writing down the ODE limit of the Markov process, as done in [20].

IV. Numerical Simulations

In this section, we conduct numerical simulations to demonstrate the competitive performance of the proposed Kaczmarz method. To begin with, we generate the subspace $\mathbf{U} \in \mathbb{R}^{n\times r}$ with i.i.d. standard Gaussian entries $\mathcal{N}(0,1)$, and $\mathbf{a}_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, with i.i.d. standard Gaussian entries $\mathcal{N}(0,1)$. The reconstruction error at each iteration is measured by the normalized mean squared error (NMSE), defined as $10\log_{10}\left(d(\mathbf{U}_k, \mathbf{U})/\|\mathbf{U}\|_{\mathrm{F}}^2\right)$ in dB. All the computations are performed on a MacBook Air with a 1.7 GHz Intel Core i7 processor and 8GB memory.

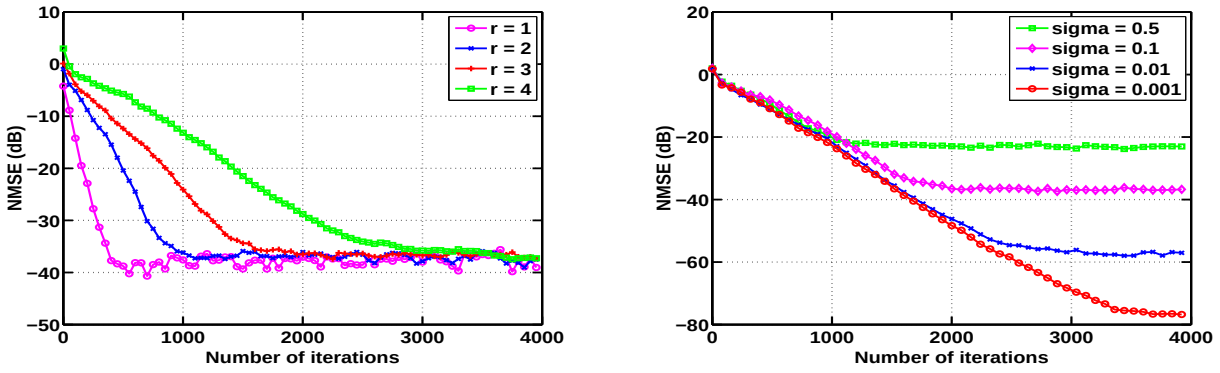
We first consider the online setting where each measurement is only accessed once by the Kaczmarz method besides the initialization. Let $n = 50$ and $m = 2500$. Fig. 1 shows the NMSE with respect to the number of iterations using either the spectral initialization or the random initialization when $r = 3$. It can be seen that the error decreases linearly as we increase the number of iterations, and using the spectral initialization improves the performance over the random initialization. For the rest of the simulations, we all adopt spectral initialization.

We now consider the performance of the proposed algorithm when we cyclically re-use all the measurements. Fig. 2 shows the NMSE with respect to the number of iterations using 5 passes of all measurements, when $n = 50$ $m = 800$, and $r = 1, 2, 3, 4$. It can be seen that the errors increase gracefully as we increase the rank.

Next, we examine the performance when the measurements are further corrupted by noise. Let the noisy measurements be

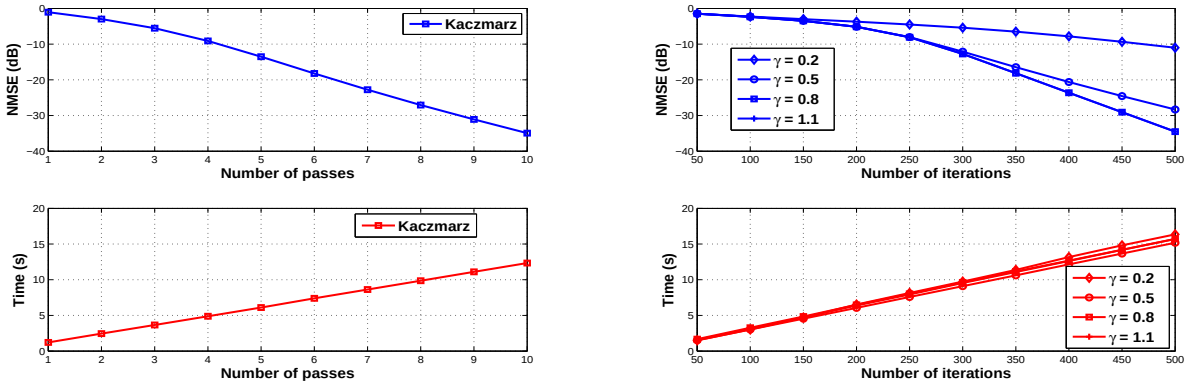
$$y_i = \left\|\mathbf{U}^T\mathbf{a}_i\right\|_2 + n_i, \quad i = 1, \ldots, m,$$

where $n_i \sim \mathcal{N}(0, \sigma^2)$ is drawn independently. Fig. 3 shows the NMSE with respect to the number of iterations under the



(a) (b)

Fig. 3. The NMSE with respect to the number of iterations with 5 passes of all measurements using spectral initialization, when $n = 50$ and $m = 800$ for noisy data: (a) fix $\sigma = 0.1$ for various ranks; (b) fix $r = 3$ for various noise levels.



(a) Kaczmarz (b) WF

Fig. 4. The NMSE and running time with respect to the number of passes or iterations using spectral initialization for (a) the proposed Kaczmarz method and (b) the WF algorithm, when $n = 1500$, $m = 10000$ and $r = 3$.

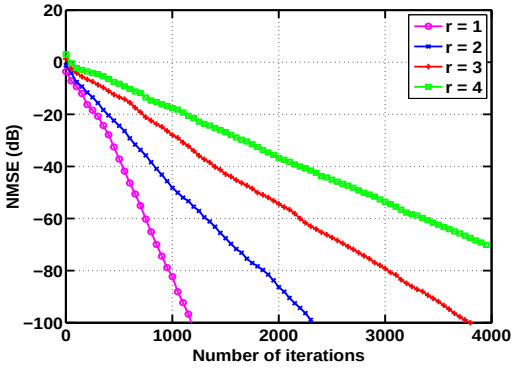


Fig. 2. The NMSE with respect to the number of iterations with 5 passes of all measurements using spectral initialization, at various ranks, when $n = 50$ and $m = 800$ for noise-free data.

same setting as Fig. 2, when (a) $\sigma = 0.1$ for various ranks, and (b) $r = 3$ for various noise levels. It can be seen that the proposed algorithm continues to perform in a stable manner in the presence of noise.

Finally, we compare the proposed algorithm with state-of-the-art, the WF algorithm [12]–[14], whose update rule can be written as

$$\mathbf{U}_k^{\mathrm{WF}} = \mathbf{U}_{k-1}^{\mathrm{WF}} + \frac{\mu_k^{\mathrm{WF}}}{m} \sum_{i=1}^{m} (z_i - \|(\mathbf{U}_k^{\mathrm{WF}})^T \mathbf{a}_i\|_2^2) \mathbf{a}_i \mathbf{a}_i^T \mathbf{U}_{k-1}^{\mathrm{WF}},$$

where $\mu_k^{\mathrm{WF}} = \min(1 - e^{-k/330}, \gamma)/\|\mathbf{U}_0^{\mathrm{WF}}\|_{\mathrm{F}}^2$ [12] with a varying step-size γ. Fig. 4 shows the NMSE and running time with respect to the number of passes or iterations using the same spectral initialization, for (a) the proposed Kaczmarz method and (b) the WF algorithm with different γ, when $n = 1500$, $m = 10000$ and $r = 3$. It can be seen that, for this example, the Kaczmarz method achieves a lower error with a faster running time, without requiring tuning parameters such as step sizes, making it a compelling alternative for large-scale problems.

V. Conclusions

In this paper, we have developed a fast iterative algorithm for estimating low-rank positive-semidefinite matrices from symmetric rank-one measurements based on generalizing the Kaczmarz method to solving a particular class of quadratic equations. It is straightforward to extend our discussions to the complex case. The proposed algorithm is extremely fast and compares favorably with the state-of-the-art algorithm based on gradient descent. On the other hand, the proposed Kaczmarz method can be regarded as applying stochastic gradient descent [22] to the loss function considered in [23].

REFERENCES

[1] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[2] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *Information Theory, IEEE Transactions on*, vol. 56, no. 6, pp. 2980–2998, 2010.

[3] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *Information Theory, IEEE Transactions on*, vol. 57, no. 4, pp. 2342–2359, 2011.

[4] K. Lee and Y. Bresler, "Admira: Atomic decomposition for minimum rank approximation," *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4402–4416, 2010.

[5] L. Tian, J. Lee, S. B. Oh, and G. Barbastathis, "Experimental compressive phase space tomography," *Optics express*, vol. 20, no. 8, pp. 8296–8308, 2012.

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

[7] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, p. 199, 2013.

[8] M. Fickus and D. G. Mixon, "Projection retrieval: Theory and algorithms," in *Sampling Theory and Applications (SampTA), 2015 International Conference on*. IEEE, 2015, pp. 183–186.

[9] 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.

[10] R. Kueng, H. Rauhut, and U. Terstiege, "Low rank matrix recovery from rank one measurements," *Applied and Computational Harmonic Analysis*, 2015.

[11] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Advances in Neural Information Processing Systems*, 2013, pp. 2796–2804.

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

[13] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *arXiv preprint arXiv:1505.05114*, 2015.

[14] C. D. White, R. Ward, and S. Sanghavi, "The local convexity of solving quadratic equations," *arXiv preprint arXiv:1506.07868*, 2015.

[15] S. Kaczmarz, "Angenäherte auflösung von systemen linearer gleichungen," *Bulletin International de l'Academie Polonaise des Sciences et des Lettres*, vol. 35, pp. 355–357, 1937.

[16] T. Strohmer and R. Vershynin, "A randomized kaczmarz algorithm with exponential convergence," *Journal of Fourier Analysis and Applications*, vol. 15, no. 2, pp. 262–278, 2009.

[17] D. Needell, "Randomized kaczmarz solver for noisy linear systems," *BIT Numerical Mathematics*, vol. 50, no. 2, pp. 395–403, 2010.

[18] D. Needell and J. A. Tropp, "Paved with good intentions: analysis of a randomized block kaczmarz method," *Linear Algebra and its Applications*, vol. 441, pp. 199–221, 2014.

[19] K. Wei, "Solving systems of phaseless equations via kaczmarz methods: a proof of concept study," *Inverse Problems*, vol. 31, no. 12, p. 125008, 2015.

[20] G. Li, Y. Gu, and Y. M. Lu, "Phase retrieval using iterative projections: Dynamics in the large systems limit," *Proc. Annual Allerton Conference on Communication, Control, and Computing*, 2015.

[21] L. Mirsky, "A trace inequality of john von neumann," *Monatshefte für Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

[22] D. Needell, R. Ward, and N. Srebro, "Stochastic gradient descent, weighted sampling, and the randomized kaczmarz algorithm," in *Advances in Neural Information Processing Systems*, 2014, pp. 1017–1025.

[23] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic systems of equations," *arXiv preprint arXiv:1605.07719*, 2016.