# Phase retrieval via randomized block Kaczmarz by averaging with heavy ball momentum

Qiyang Duan [1, a], Lu Zhang [1, b], Hui Zhang [1, c] and Hongxia Wang [1, d]

[1]National University of Defense Technology, Department of Mathematics, Changsha 410073, Hunan, China.

[a] duanqiyang@nudt.edu.cn, [b] zhanglu21@nudt.edu.cn, [c] h.zhang1984@163.com,

[d] wanghongxia@nudt.edu.cn (corresponding author)

**Abstract.** We propose a variant of the randomized Kaczmarz method for solving phase retrieval problems called randomized block Kaczmarz with heavy ball momentum (RBK-HB). It achieves effective acceleration compared to the Kaczmarz methods by combining block and heavy ball momentum techniques. In the theoretical part, by assuming that the loss function is strongly convex near the true solution, the RBK-HB method converges linearly with high probability. Numerical experiments show that compared with the Kaczmarz methods, the RBK-HB method is less sensitive to the initial point, the number of measurements required for successful recovery is less and has a faster convergence rate.

**Keywords:** Phase retrieval; randomized block Kaczmarz; heavy ball momentum.

## 1. Introduction

Phase retrieval, which aims to recover a vector from its intensity measurements, i.e., to solve a system of phaseless equations:

$$y_r = |\langle a_r, x\rangle|^2, r = 1, \cdots, m, \#(\quad)$$

Where $x \in C^n$ is the signal to be recovered and $a_r \in C^n$ is the measurement vector. Let $A \in C^{m\times n}$ be a matrix whose rows are $\{a_r^*\}_{1\le r\le m}$ ($[\cdot]^*$ is a conjugate transpose of a vector) and $y = (y_1, ..., y_m)^T$, the phase retrieval problem can be formulated as solving $y = |Ax|^2$.

The Kaczmarz method is first proposed by Kaczmarz for solving linear systems $Ax = y$ [1]. In the k-th iteration, the new estimate $x_{k+1}$ is obtained by projecting current estimate $x_k$ onto the hyperplane $\{x: \langle a_r, x\rangle = y_r\}$ as

$$x_{k+1} = x_k + \frac{y_r - \langle a_r, x\rangle}{\|a_r\|_2^2} a_r.$$

In [2], the Kaczmarz method is used to solve the phase retrieval problem for the first time. It keeps the orthogonal projection but considers the different hyperplane $\left\{x: \langle a_r, x\rangle = \sqrt{y_r} e^{i\theta_k}\right\}$, Where $\theta_k = \angle\langle a_r, x\rangle$, i.e. the image phase of the solution is approximated by that of the current estimate.

We propose a variant of the Kaczmarz method for phase retrieval called RBK-HB, modified to include the block technique and a heavy ball momentum term (see Algorithm 1).

## 2. Randomized block Kaczmarz with heavy ball momentum method

In the k-th iteration, instead of using an individual row, we select $\eta$ ($\eta \ge 1$) rows randomly and project the current estimate onto each selected row, which improves the utilization of information. We assume measurement vectors $a_i \in C^n$ are normalized, and consider the following iterative format:

$$\tilde{x}_k = x_k - \frac{\alpha}{\eta} \sum_{i\in\gamma_k} \left(1 - \frac{\sqrt{y_i}}{|a_i^* x_k|}\right) a_i\, a_i^* x_k, \#(1)$$

where the elements in index set $\gamma_k$ are chosen uniformly from $\{1, \cdots, m\}$ and $|\gamma_k| = \eta$, $\alpha$ is a relaxation factor generally takes 1.

Furthermore, we leverage heavy ball momentum term to obtain a probable acceleration:

$$x_{k+1} = \tilde{x}_k + \beta(x_k - x_{k-1}), \#(2)$$

Where β is the parameter of the heavy ball momentum term. We summarize the above in Algorithm 1.

**Algorithm 1** Randomized block Kaczmarz with heavy ball momentum (RBK-HB)

**Input**: $x_0, \eta, \alpha, \beta$

**For** $k = 0,1,\dots,K$ **do**

select a block of $A$, denoted by $A_{\gamma_k}$, $|\gamma_k| = \eta$ *and* $\gamma_k \sim P$

$\theta_i^k = \angle\langle a_r, x\rangle, i \in \gamma_k$

$X_{k+1} = x_k + \frac{\alpha}{\eta}\sum_{i\in\gamma_k}\left(\frac{\sqrt{y_i}e^{i\theta_i^k} - \langle a_i, x_k\rangle}{\|a_i\|_2^2}\right)a_i + \beta(x_k - x_{k-1})$

**End for**

**Output**: $x_K$

The update (2) of $x_k$ can be considered as an SGD update of step size $\alpha/2$ for the following loss function:

$$f(x) = \frac{1}{m}\sum_{i=1}^{m}(|a_i^* x| - \sqrt{y_i})^2 . \#(3)$$

Its one-sided directional derivative at x along the direction v is given by

$$f_v'(x) = \lim_{t\to 0^+}\frac{f(x+tv) - f(x)}{t} = \frac{1}{m}\sum_{i=1}^{m}(1 - \frac{\sqrt{y_i}}{|a_i^* x_k|})(a_i^* v x^* a_i + a_i^* x v^* a_i).$$

The loss function (4) and its local regularity property has been studied in [3,4], they assume

$$f_{x-z}'(x) \ge \frac{\mu}{2}\left\|f'(x)\right\|^2 + \frac{\lambda}{2}\|x - z\|^2$$

and only obtain the results of linear convergence in a real-valued setting.

To analyze the theoretical results in a complex setting, we assume f satisfies the local restricted strong convexity, that is:

$$f(x) + f_{z-x}'(x) + \frac{L}{n}\|x - z\|^2 \le f(z), x \in B(z, c_0), \#(4)$$

where $B(z, c_0) := \{x \mid \|x - z\| \le c_0\}$ and $c_0$ is a positive constant. This property essentially states that the gradient of the function is well-behaved, which ensures the point in $B(z, c_0)$ converges linearly to the true solution with a high probability. Similar assumptions can be found in [5,6], but our results are more general (when $\eta = 1$, $\beta = 0$, their result is a special case of ours).

## 3. Main Convergence Result

Assuming (5) holds, we conclude that the RBK-HB method converges linearly with high probability.

**Theorem 1** Assume $\|a_i\| = 1$ for all $1 \le i \le m$, x is the true solution and (5) holds, the RBK-HB method uniformly selects a block γ in each iterate, and the initilization $x_0$ such that $\|x_0 - x\| \le c_0\sqrt{\delta_1}$ for some $0 \le \delta_1 \le \frac{1}{2}$. If $0 < \beta < \min\{\frac{L}{4n\eta}, \frac{1}{2}\}$, fix $\varepsilon > 0$ and some $q \in (0,1)$ then

$$\|x_k - x\|^2 \le \varepsilon\|x_0 - x\|^2$$

holds with probability at least $1 - \delta_1 - \frac{2q^k(1+\delta)}{\varepsilon}$, where $q = \frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2}$, $a_1 = \frac{1 - \frac{L}{\eta n} + \beta}{1 - 2\beta}$, $a_2 = \frac{\beta}{1 - 2\beta}$.

First, we present lemma 9 in [7], which we will use in our proof.

**Lemma 1** Let $\{F_k\}_{k\geq 0}$ be nonnegative real numbers. Fix $F_1 = F_0 \geq 0$ and assume that $\{F_k\}_{k\geq 0}$ satisfies:

$$F_{k+1} \leq a_1 F_k + a_2 F_{k-1}, \forall\, k \geq 1,$$

where $a_2 \geq 0, a_1 + a_2 < 1$ and at least one of the coefficients $a_1, a_2$ is positive. Then the following inequality holds for all $k \geq 1$:

$$F_{k+1} \leq q^k (1+\delta) F_0,$$

where $q = \frac{a_1 + \sqrt{a_1{}^2 + 4a_2}}{2}$, $\delta = q - a_1 \geq 0$ and $q \geq a_1 + a_2$.

Now we turn to the proof of Theorem 1.

***Proof 3.1.*** According to (2), we have

$$\begin{aligned}
&\|\tilde{x}_k - x\|^2 \\
&= \|x_k - x - \frac{1}{\eta}\sum_{i\in\gamma_k}(1 - \frac{\sqrt{y_i}}{|a_i^* x_k|}) a_i\, a_i^* x_k\|^2 \\
&\leq \|x_k - x\|^2 + \frac{1}{\eta^2}\sum_{i\in\gamma_k}\|\left(1 - \frac{\sqrt{y_i}}{|a_i^* x_k|}\right) a_i a_i^* x_k\|^2 - \frac{1}{\eta}\sum_{i\in\gamma_k} 2\mathrm{Re}[(1 - \frac{\sqrt{y_i}}{|a_i^* x_k|}) a_i\, a_i^* x_k (x_k - x)] \qquad (1)
\end{aligned}$$

Taking the expectation of (6), we have

$$\begin{aligned}
&E\|\tilde{x}_k - x\|^2 \\
&= \|x_k - x\|^2 + \frac{1}{\eta^2 m}\sum_{i=1}^{m}(|a_i^* x_k| - \sqrt{y_i})^2 + \frac{1}{\eta m}\sum_{i=1}^{m} 2\mathrm{Re}[(1 - \frac{\sqrt{y_i}}{|a_i^* x_k|})\, a_i a_i^* x_k (x - x_k)] \\
&= \|x_k - x\|^2 + \frac{1}{\eta^2} f(x_k) + \frac{1}{\eta} f'_{x - x_k}(x_k) \\
&\leq \left(1 - \frac{L}{\eta n}\right)\|x_k - x\|^2, \qquad (2)
\end{aligned}$$

Where $2\mathrm{Re}(x) = x + x^*$, and the last inequality applies the assumption (5) with $f(x) = 0$ and $\eta \geq 1$. Together with $x_{k+1} = \tilde{x}_k + \beta(x_k - x_{k-1})$ we get

$$\begin{aligned}
&E\|\tilde{x}_k - x\|^2 \\
&= E\|x_{k+1} - x - \beta(x_k - x) + \beta(x_{k-1} - x)\|^2 \\
&= E\|x_{k+1} - x\|^2 - 2\beta E\left[\mathrm{Re}\left((x_{k+1} - x)^*(x_k - x)\right)\right] + 2\beta E\left[\mathrm{Re}\left((x_{k+1} - x)^*(x_{k-1} - x)\right)\right] \\
&\quad + \beta^2\|x_k - x\|^2 + \beta^2\|x_{k-1} - x\|^2 - 2\beta^2\mathrm{Re}\left((x_k - x)^*(x_{k-1} - x)\right) \\
&\geq (1 - 2\beta)E\|x_{k+1} - x\|^2 - \beta\|x_k - x\|^2 - \beta\|x_{k-1} - x\|^2. \qquad (3)
\end{aligned}$$

The last inequality uses the fact that

$$|2\mathrm{Re}(x^* y)| \leq \|x\|^2 + \|y\|^2, \forall x, y \in C^n. \qquad (4)$$

Combining (7) and (8) yields

$$E\|x_{k+1} - x\|^2 \leq \frac{1 - \frac{L}{\eta n} + \beta}{1 - 2\beta}\|x_k - x\|^2 + \frac{\beta}{1 - 2\beta}\|x_{k-1} - x\|^2, \qquad (5)$$

when $1 - 2\beta > 0$.

Finally, we apply Lemma 1, where the two coefficients are given by $a_1 = \frac{1 - \frac{L}{\eta n} + \beta}{1 - 2\beta}$ *and* $a_2 = \frac{\beta}{1 - 2\beta} > 0$. Since we require $a_2 \geq 0, a_1 + a_2 < 1$ and at least one of the coefficients $a_1, a_2$ is positive, thus the assumptions for Lemma 1 hold if

$$0 < \beta < \min\{\frac{L}{4n\eta}, \frac{1}{2}\}.$$

Notice that we give an upbound of $\beta$ by inequality scaling (9), but $\beta$ could be wider by adjusting the inequality (9). Thus, we obtain

$$E\|x_k - x\|^2 \leq q^k (1+\delta)\|x_0 - x\|^2,$$

Where $q = \frac{a_1 + \sqrt{a_1{}^2 + 4a_2}}{2}$, $\delta = q - a_1$ and $1 > q \geq a_1 + a_2$. It follows from $0 < q < 1$ that the iterates generated from the RBK-HB method converge linearly in expectation.

The rest of the proof follows from Theorem 3.1 and Corollary 3.2 in [8]. Fix $\varepsilon > 0, 0 < \delta_1 \le \frac{1}{2}$. Let $\tau = \min\{k: \|x_k - x\| \ge c_0\}$, we have $P(\tau < \infty) \le \left(c_0\sqrt{\delta_1}/c_0\right)^2 = \delta_1$, and $P(\|x_k - x\|^2 \le \varepsilon\|x_0 - x\|^2) \ge 1 - \delta_1 - \frac{2q^k(1+\delta)}{\varepsilon}$. This completes the proof.

## 4. Numerical experiments

In this section, we present some numerical experiments to test our method. All experiments are performed with MATLAB on a personal computer with 2.80-GHZ CPU(Intel(R) Core(TM) i7-1165G7) and 16-GB memory.

The Measured complex vectors $\hat{x}$ are generated from the Gaussian distribution, that is, $\hat{x} \sim N(0, I_n) + iN(0, I_n)$. The algorithm is tested with the Gaussian model with entries of A drawn i.i.d. from $N\left(0, \frac{1}{2}\right) + iN\left(0, \frac{1}{2}\right)$, and the measurements $y = |A\hat{x}|^2$. We set $\alpha = 1$, $n = 128$ and $m = [\delta \times n]$, where $\delta \ge 2$. The initial point is set as $x_0 = \sqrt{\frac{\sum_{r=1}^{m} y_r}{m}} z$, where z is the unit leading eigenvector of $\sum_{r=1}^{m} y_r\, a_r a_r^*$ suggested by [9]. The k-th relative error is defined by

$$e(k) = \frac{\mathrm{dist}(x_k, \hat{x})}{\|\hat{x}\|},$$

Where $\mathrm{dist}(x_k, \hat{x}) = \min_{\theta \in R}\left\|x_k - \hat{x}e^{i\theta}\right\|$.

### 4.1 Convergence Rate in Different Settings.

To explore the influence of η and β on the performance of our method, we set $\delta = 4$. Let $\eta \in \{1,8\}$, $\beta \in \{0, 0.5, 0.9\}$ and the maximum iteration is 100000. When $\eta = 1$ and $\beta = 0$ the RBK-HB method is the Kaczmarz method. We find that when η takes 1, as long as β is greater than 0.5, the algorithm will not converge. The results are shown in Figure 1.

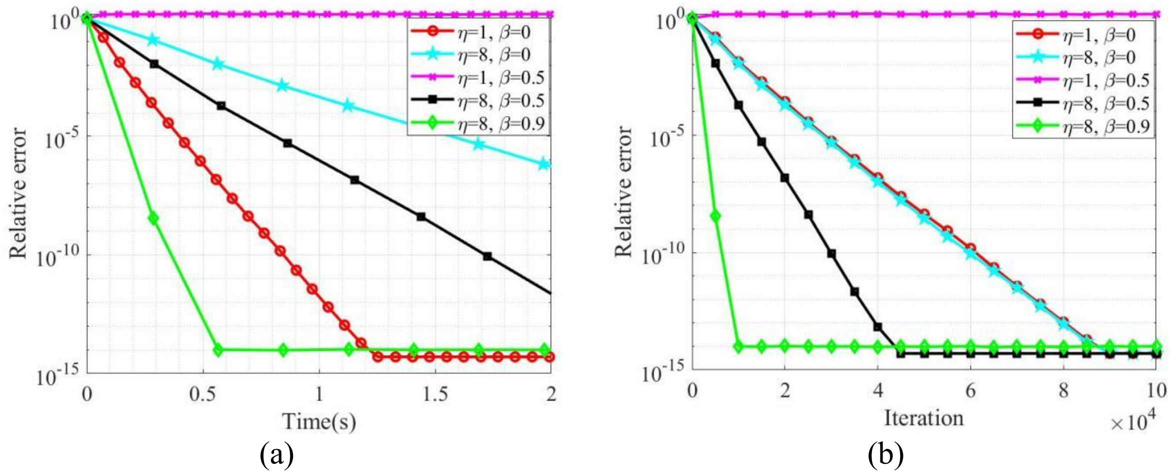
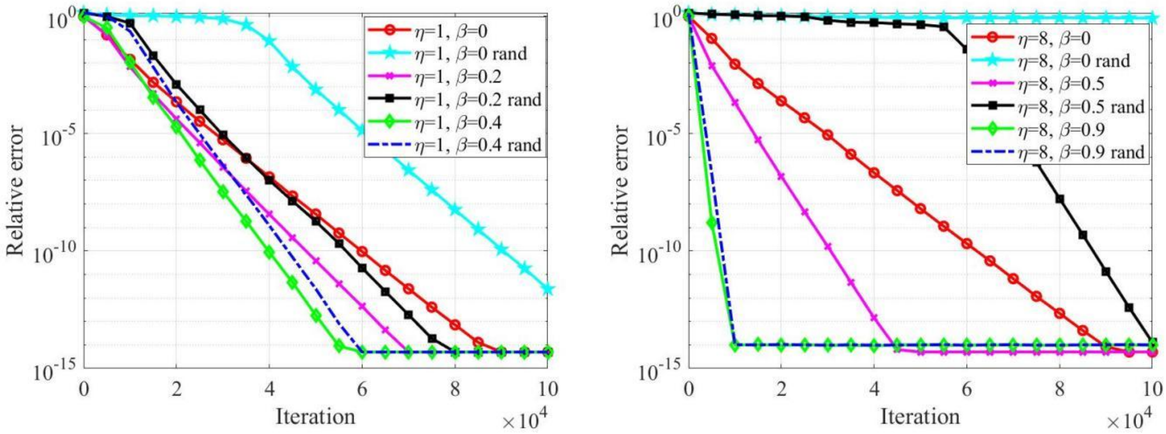



Fig. 1 Convergence rate for different η and β when n=128 and m=512. (a) The relationship between relative error and running time; (b) The relationship between relative error and iteration.

From Figure 1, we can see that the main source of acceleration is the heavy ball momentum term since the two block size schemes have the same decrease in each iteration and the scheme of large block takes longer time when β takes 0. This is because η does not affect the step size of the Kaczmarz (SGD) method, and the computational cost of each iteration increases when η gets larger. However, blocking is necessary because only increasing β will lead to non-convergence of the algorithm. When η takes 1 and β takes 0.5, the algorithm does not converge, while when η takes 8, β is able to take 0.9 to achieve a half-time improvement compared with the Kaczmarz method.

**4.2 Sensitivity to initial points.**

For non-convex optimization, the initialization procedure is very important and a good initial point can prevent convergence to a local minimum. In this section, we test the sensitivity of the RBK-HB method to initial points by using two different initializations, one is the refined initial point mentioned at the beginning of this section, and the other is randomly generated according to the standard Gaussian distribution. The results are shown in Figure 2.




(a) (b)

Fig. 2 Convergence rates under two different initializations. (a) The effect of β on the sensitivity of the RBK-HB method to the initial point when $\eta = 1$; (b) The effect of β on the sensitivity of the RBK-HB method to the initial point when $\eta = 8$.

From Figure 2, the RBK-HB method performs better for refined initialization than random initialization. However, this difference will gradually decrease with the increase of β. When β takes 0.9, the difference between the two initializations is very small, which reminds us that β can be appropriately increased to reduce the sensitivity of RBK-HB to the initial point.

**4.3 Robustness to noise.**

In practical applications, measurement noise is inevitable. In this subsection, we will test our method's ability to resist additive noise which is modelled as follows:

$$\sqrt{y_r} = |\langle a_r, x\rangle + \varepsilon_r|, r = 1, \cdots, m,$$

where $\varepsilon_r$ is the Gaussian noise. The signal noise ratio (SNR) of the vector to be recovered is set to 0dB to 60dB and the maximum iteration is 30000. The results are shown in Figure 3.

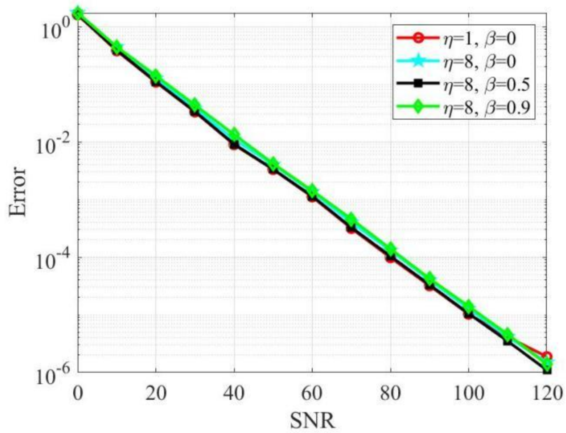


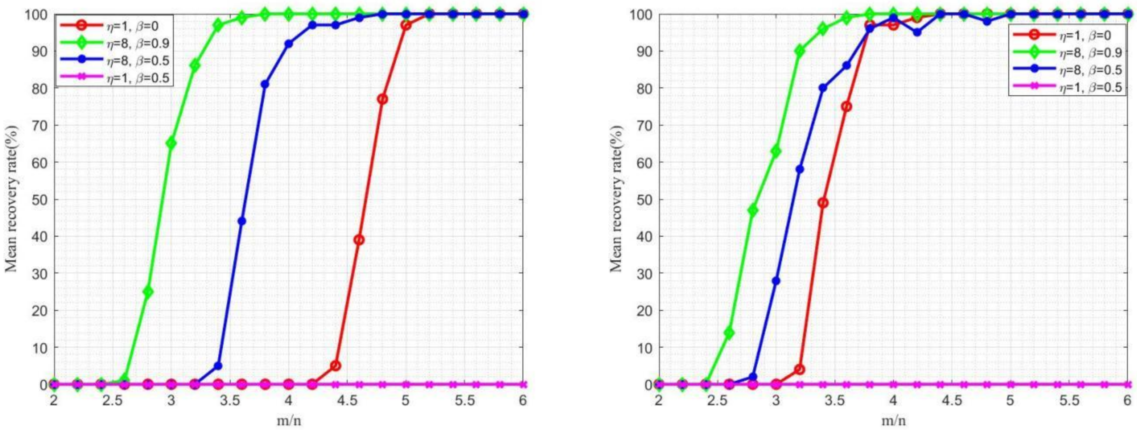

Fig. 3 The relationship between SNR and relative error at different Settings.

From Figure 3, we can find that the relative error decreases steadily with the increase of SNR under each setting. This proves that our method is robust to additive noise.

**4.4 The Recovery Probability.**

The number of measurements $m$ has a crucial influence on the phase retrieval algorithm. In this subsection, we test the relationship between the mean recovery rate and the number of measurements under different settings.

We set $\delta = 2:0.2:6$ where ':' is the Matlab notation indicating that the interval is 0.2. Let $\eta \in \{1,4\}$, $\beta \in \{0,0.2,0.5\}$and the maximum iteration be 20000 and 50000 respectively. An experiment is considered successful if the relative error is less than $10^{-5}$ within the maximum iteration. We randomly generate 100 measurement matrices and signals under each setting and repeat the experiment 100 times, the number of successful experiments is recorded as the mean recovery rate. The results are shown in Figure 4.




(a) (b)

Fig. 4 The relationship between the mean recovery rate and the number of measurements under different settings. (a) The mean recovery rate of 20000 iterations; (b) The mean recovery rate of 50000 iterations.

From Figure 4, we can find that the number of measurements required for the RBK-HB method to successfully recover a vector is smaller than the Kaczmarz method. Furthermore, this difference is more obvious when the maximum iteration is small.

## Funding

This work was supported by National Natural Science Foundation of China (61977065) and National Key Research and Development Program (2020YFA0713504).

## References

[1] S. Karczmarz, "Angenaherte auflosung von systemen linearer glei-chungen," Bull. Int. Acad. Pol. Sic. Let., Cl. Sci. Math. Nat. pp. 355–357 (1937).

[2] K. Wei, "Solving systems of phaseless equations via kaczmarz methods: A proof of concept study," Inverse Probl. 31, 125008 (2015).

[3] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," IEEE Trans. on Inf. Theory 64, 773–794 (2017).

[4] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic system of equations," Adv. Neural Inf. Process. Syst. 29 (2016).

[5] M. Huang and Y. Wang, "Linear convergence of randomized kaczmarz method for solving complex-valued phaseless equations," SIAM J. on Imaging Sci. 15, 989–1016 (2022).

[6] T. Zhang and Y. Feng, "Phase retrieval of complex-valued objects via a randomized kaczmarz method," Inf. Inference: A J. IMA 11, 823–843 (2022).

[7] N. Loizou and P. Richtárik, "Momentum and stochastic momentum for stochastic gradient, newton, proximal point and subspace descent methods," Comput. Optim. Appl. 77, 653–710 (2020).

[8] Y. S. Tan and R. Vershynin, “Phase retrieval via randomized kaczmarz: theoretical guarantees,” Inf. Inference: A J. IMA 8, 97–123 (2019).

[9] E. J. Candes, X. Li, and M. Soltanolkotabi, “Phase retrieval via wirtinger flow: Theory and algorithms,” IEEE Trans. on Inf. Theory 61, 1985–2007 (2015).