

# An accelerated randomized extended Kaczmarz algorithm

Xu Xiang<sup>1,\*</sup>, Xu Liu<sup>1</sup>, Wentang Tan<sup>1</sup> and Xiang Dai<sup>1</sup>

<sup>1</sup>Southwest Electronics and Telecommunication Technology Research Institute, Chengdu, Sichuan, China, 610041

\* E-mail address: xiangxu1402@163.com

**Abstract.** The randomized Kaczmarz(RK) is a useful algorithm for solving consistent linear system  $\mathbf{Ax} = \mathbf{b}$  ( $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}$ ). It was proved that for inconsistent linear system, with randomized orthogonal projection, the randomized extended Kaczmarz(REK) method converges with an expected exponential rate. We describe an accelerated randomized extended Kaczmarz algorithm(AREK) with Nesterov's accelerated procedure. The analysis shows that AREK converges better than REK when A is dense and the smallest singular value of  $\mathbf{A}^T \mathbf{A}$  is small.

## 1. Introduction

The Kaczmarz method[1] is a popular algorithm for solving overdetermined linear systems and has numerous applications from tomography to image processing[2, 3]. This method sweeps through the rows of A in a cyclic manner, projecting in each substep the last iterate orthogonally onto the solution hyperplane of  $\langle \mathbf{a}_i, \mathbf{x} \rangle = \mathbf{b}_i$ ; then the  $(k+1)^{\text{th}}$  estimate  $\mathbf{x}_{k+1}$  is defined by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\mathbf{b}_i - \langle \mathbf{a}_i, \mathbf{x}_k \rangle}{\|\mathbf{a}_i\|_2^2} \mathbf{a}_i, \quad (1)$$

where  $i = k \bmod m + 1$ ,  $\mathbf{a}_i$  is the  $i^{\text{th}}$  row of A.

T. Strohmer and R. Vershynin [4, 5] propose at each iteration to randomly select a row of A with probability proportional to Euclidean norm of the row. This method is the randomized Kaczmarz(RK) which can thus be described by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\mathbf{b}[p(i)] - \langle \mathbf{a}_{p(i)}, \mathbf{x}_k \rangle}{\|\mathbf{a}_{p(i)}\|_2^2} \mathbf{a}_{p(i)}, \quad (2)$$

where  $p(i)$  takes values in  $\{1, \dots, m\}$  with probabilities  $\frac{\|\mathbf{a}_{p(i)}\|_2^2}{\|\mathbf{A}\|_F^2}$ . This randomized version of the Kaczmarz method provides clear advantages over the standard method in many cases. And in [5], using the selection above, they were able to provide a proof for the expected rate of convergence,

$$\mathbb{E} \|\mathbf{x}_k - \mathbf{x}^*\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \cdot \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \quad (3)$$



where  $x_0$  denotes the initial approximation, assuming that  $\mathbf{Ax} = \mathbf{b}$  has a solution and denote  $\mathbf{x}^* := \mathbf{A}^{\dagger}\mathbf{b}$ ,

$\mathbf{R} = \|\mathbf{A}^{-1}\|_2^2 \|\mathbf{A}\|_F^2$ ,  $\|\mathbf{A}^{-1}\|_2$  is the smallest constant  $M$  such that the inequality  $\|\mathbf{Ax}\|_2 \geq \frac{1}{M}\|\mathbf{x}\|_2$  holds for all vectors  $\mathbf{x}$ . The RK method addresses the significant disadvantage of the original Kaczmarz algorithm. It is easy to analyze the convergence rate for the RK method, while the original algorithm converge very slow when the data order is poor.

In 2010, D. Needell [6] analyzed the case where the system  $\mathbf{Ax} = \mathbf{b}$  is corrupted by noise, and considered the system  $\mathbf{Ax} = \mathbf{b} + \mathbf{r}$  where  $\mathbf{r}$  is an arbitrary error vector. In the noisy version, she obtained the RK method reaches an error threshold dependent on the matrix  $\mathbf{A}$  with the same rate as in the error-free case,

$$\mathbb{E}\|\mathbf{x}_k^* - \mathbf{x}^*\|_2 \leq \left(1 - \frac{1}{\mathbf{R}}\right)^{\frac{k}{2}} \|\mathbf{x}_0 - \mathbf{x}^*\|_2 + \sqrt{\mathbf{R}}\gamma, \quad (4)$$

where  $\gamma = \max_i |r_i| / \|\mathbf{a}_i\|_2$ ,  $x_k^*$  denote the  $k^{\text{th}}$  iterate of noisy randomized Kaczmarz. Later, in [7], D. Needell and Y. C. Eldar utilized the Johnson-Lindenstrauss [7] dimension reduction technique to keep the runtime on the same order as the original RK version and proposed randomized Kaczmarz via Johnson-Lindenstrauss (RKJL) and improved the convergence rate of the RK method.

In [8], A. Zouzias and N. M. Freris presented the randomized extended Kaczmarz for solving inconsistent linear system  $\mathbf{Ax} = \mathbf{b} + \mathbf{r}$ . By a randomized approximation orthogonal projection, they approximately computed linear equation of  $\mathbf{Ax} = \mathbf{b}_{\mathcal{R}(\mathbf{A})}$ , where  $\mathcal{R}(\mathbf{A}) := \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\}$ . Essentially, the RK algorithm is applied twice. The convergence rate of this method is

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}^*\|_2^2 \leq \left(1 - \frac{1}{\mathbf{R}}\right)^{\lfloor \frac{k}{2} \rfloor} \left(1 + 2\mathbf{R}_1^2\right) \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \quad (5)$$

Let  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\text{rank}(\mathbf{A})}$  be the non-zero singular values of  $\mathbf{A}$ , where  $\mathbf{R}_1^2$  denotes  $\sigma_1^4 / \sigma_{\text{rank}(\mathbf{A})}^2$ . This method can be considered as a randomized variant of the extended Kaczmarz method proposed by C. Popa [9].

In [10], J. Liu and S. J. Wright proposed an accelerated randomized (ARK) algorithm with better convergence than the standard RK method on ill conditioned problems. The ARK method starts at a significant disadvantage in the sparse setting, so by Nesterov's accelerated procedure [11], they improved the iteration of the standard RK method and proved, if matrix  $\mathbf{A}$  is dense and the minimum singular value of  $\mathbf{A}^T \mathbf{A}$  is small,

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}^*\|_2^2 \leq \frac{4m^2 \|\mathbf{x}_0 - \mathbf{x}^*\|_{(\mathbf{A}^T \mathbf{A})^+}^2}{k^2}, \quad (6)$$

where  $\|\mathbf{x}_0 - \mathbf{x}^*\|_{(\mathbf{A}^T \mathbf{A})^+}^2$  is  $\sqrt{\text{trace}((\mathbf{x}_0 - \mathbf{x}^*)^T (\mathbf{A}^T \mathbf{A})^+ (\mathbf{x}_0 - \mathbf{x}^*))}$ ,  $(\mathbf{A}^T \mathbf{A})$  is the Moore-Penrose pseudoinverse of  $\mathbf{A}^T \mathbf{A}$ . Motivated by the idea of Nesterov's acceleration, unlike the asynchronous parallel randomized Kaczmarz algorithm, we apply Nesterov's accelerated procedure in the REK method if  $\mathbf{A}$  is dense and the minimum singular value of  $\mathbf{A}^T \mathbf{A}$  is small, and propose an accelerated randomized extended Kaczmarz algorithm (AREK). In this paper, under the condition that  $\mathbf{A}$  is dense and the minimum singular value of  $\mathbf{A}^T \mathbf{A}$  is small, we prove that the convergence rate of the AREK method is better than the REK method, and illustrate the computational results of our approach.

## 2. Algorithm

Both theoretical and numerical researches have shown that the RK algorithm provides very promising results. Here we mainly show an accelerated REK method works well in the case where the system is corrupted with noise. In this section we consider the consistent system  $\mathbf{Ax} = \mathbf{b}$  after an error  $\mathbf{r}$  is added to the right side,  $\mathbf{Ax} \approx \mathbf{b} + \mathbf{r}$ . In (4) with  $\mathbf{r} = \mathbf{b}_{\mathcal{R}(\mathbf{A})}$  tells us that the randomized Kaczmarz algorithm

performs well for least square problems whose least squares error is very close to zero,  $\|\mathbf{r}\|_2 \approx 0$ . First we give the notations and preliminaries.

### 2.1 Notations and preliminaries

We denote the rows and columns of  $A$  by  $\mathbf{a}_1, \dots, \mathbf{a}_m$  and  $\mathbf{a}^1, \dots, \mathbf{a}^n$ . For an integer  $m$ , let  $[m] := \{1, 2, \dots, m\}$ . Let  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\text{rank}(A)}$  be the non-zero singular values of  $A$ ,  $\sigma_1 = \sigma_{\min}$  and  $\sigma_{\text{rank}(A)} = \sigma_{\max}$ .  $\mathcal{R}(A) := \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\}$  and  $\mathcal{R}(A)^\perp$  denotes the orthogonal complement of  $\mathcal{R}(A)$ . Assume that  $\mathbf{Ax} = \mathbf{b}$  has a solution and denote  $\mathbf{x}^* := A^\dagger \mathbf{b}$ , and the first estimate  $\mathbf{x}_0 = 0$ . Given any  $\mathbf{b} \in \mathbb{R}^m$ , we can write it as  $\mathbf{b}_{\mathcal{R}(A)} + \mathbf{b}_{\mathcal{R}(A)^\perp}$ , where  $\mathbf{b}_{\mathcal{R}(A)}$  is the projection of  $\mathbf{b}$  onto  $\mathcal{R}(A)$ .  $\|A\|_F$  is the Frobenius norm of  $A$ . Given a positive semidefinite matrix  $M$ ,  $\|X\|_M$  is defined as  $\sqrt{\text{trace}(X^T M X)}$ .  $R = \|A^{-1}\|_2^2 \|A\|_F^2$ ,  $\|A^{-1}\|_2$  is the smallest constant  $M$  such that the inequality  $\|\mathbf{Ax}\|_2 \geq (1/M)\|\mathbf{x}\|_2$  holds for all vectors  $\mathbf{x}$ . We assume throughout that the rows of  $A$  are normalized,  $\|\mathbf{a}_i\|_2 = 1$ ,  $i = 1, 2, \dots, m$ .

### 2.2 Randomized extended Kaczmarz(REK)

**Table 1.** Randomized extended Kaczmarz(REK) (Algorithm 1)

<b>Require:</b>	$m \times n$ matrix $A$ , coefficient vector $\mathbf{b} \in \mathbb{R}^m$
<b>Initialization:</b>	$\mathbf{x}_0 = 0$ and $\mathbf{z}_0 = \mathbf{b}$ .
<b>Select:</b>	Pick $i_k \in [m]$ , $j_k \in [n]$ with probability $q_i := \frac{\ \mathbf{a}_i\ _2^2}{\ A\ _F^2}$ , $i \in [m]$ and $p_j := \frac{\ \mathbf{a}^j\ _2^2}{\ A\ _F^2}$ , $j \in [n]$
<b>Project:</b>	$\mathbf{z}_{k+1} = \mathbf{z}_k - \frac{\langle \mathbf{a}^{j_k}, \mathbf{z}_k \rangle}{\ \mathbf{a}^{j_k}\ _2^2} \mathbf{a}^{j_k}, \mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\mathbf{b}^{i_k} - \mathbf{z}_k^{i_k} - \langle \mathbf{x}_k, \mathbf{a}_{i_k} \rangle \mathbf{a}_{i_k}}{\ \mathbf{a}_{i_k}\ _2^2} \mathbf{a}_{i_k}$
<b>Update:</b>	Set $k = k + 1$
<b>end</b>	

A. Zouzias and N. M. Freris proposed the variant of the RK method for solving inconsistent linear system, see [8] for more details.

This approach can efficiently reduce the norm of the noisy part of  $\mathbf{b}$ , then apply the RK algorithm on a new linear system whose right hand side vector is now arbitrarily close to the column space of  $A$ ,  $\mathbf{Ax} \approx \mathbf{b}_{\mathcal{R}(A)}$ . Algorithm 1, see Table 1, showed that the RK method is applied twice by the REK method, and The convergence rate of it has been proven, it is showed in (5). Actually, A. Zouzias and N. M. Freris did not give the best result of the convergence rate of the REK method. By the method of their proof, it is easy to obtain the convergence rate of the REK algorithm,

**Theorem 1.** After  $k > 1$  iterations, let  $\mathbf{x}_k^*$  be the  $k^{\text{th}}$  iterate of the noisy REK method, then,

$$\mathbb{E} \|\mathbf{x}_k^* - \mathbf{x}^*\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 + 2 \left(1 - \frac{1}{R}\right)^{\lfloor \frac{k}{2} \rfloor} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 \frac{\sigma_{\max}^2(A)}{\sigma_{\min}^2(A)}, \quad (7)$$

where  $R = \|A^{-1}\|_2^2 \|A\|_F^2$ , and the expectation is taken over the choice of the rows in the algorithm.

We now describe the complexity of the REK method for the case of dense A. In Section 2, We assume throughout that the rows of A are normalized,  $\|\mathbf{a}_i\|_2 = 1, i = 1, 2, \dots, m$ . Then the computation of Algorithm 1 requires about  $4(m + n)$  operations per iteration.

### 2.3 Nesterov's Accelerated Randomized Kaczmarz(ARK)

**Table 2.** Accelerated Randomized Kaczmarz(ARK) (Algorithm 2)

<b>Require:</b>	$m \times n$ matrix A, coefficient vector $\mathbf{b} \in \mathbb{R}^m$
<b>Initialization:</b>	$\mathbf{v}_0 = \mathbf{x}_0 = 0$ . Choose $\gamma_k$ to be the larger root as in (8). Set $\alpha_k$ and $\beta_k$ as in (9).
<b>Select:</b>	Pick $i_k \in [m]$ with probability $q_i := \frac{\ \mathbf{a}_i\ _2^2}{\ \mathbf{A}\ _F^2}, i \in [m]$
<b>Project:</b>	$\mathbf{y}_k = \alpha_k \mathbf{v}_k + (1 - \alpha_k) \mathbf{x}_k; \mathbf{x}_{k+1} = \mathbf{y}_k - \mathbf{a}_{i_k} (\mathbf{a}_{i_k}^T \mathbf{y}_k - \mathbf{b}^{i_k}) / \ \mathbf{a}_{i_k}\ _2^2;$ $\mathbf{v}_{k+1} = \beta_k \mathbf{v}_k + (1 - \beta_k) \mathbf{y}_k - \gamma_k \mathbf{a}_{i_k} (\mathbf{a}_{i_k}^T \mathbf{y}_k - \mathbf{b}^{i_k}) / \ \mathbf{a}_{i_k}\ _2^2$
<b>Update:</b>	Set $k = k + 1$
<b>end</b>	

The ARK algorithm, see Table 2, applies Nesterov's accelerated procedure [11] to improve the standard RK method. If we apply  $\mathbf{Ax} = \mathbf{b}$  to  $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2$ , gradient descent sets  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \theta_k \nabla f(\mathbf{x}_k)$ , where  $\nabla f(\mathbf{x}_k)$  is the objective gradient and  $\theta_k$  is the stepsize. Nesterov's accelerated procedure introduces two sequences  $\{\mathbf{y}_k\}$  and  $\{\mathbf{v}_k\}$  and obtain the following iterative scheme:  $\mathbf{y}_k \leftarrow \alpha_k \mathbf{v}_k + (1 - \alpha_k) \mathbf{x}_k; \mathbf{x}_{k+1} \leftarrow \mathbf{y}_k - \theta_k \nabla f(\mathbf{y}_k); \mathbf{v}_{k+1} \leftarrow \beta_k \mathbf{v}_k + (1 - \beta_k) \mathbf{y}_k + \gamma_k \nabla f(\mathbf{y}_k)$ . With appropriate choice of  $\alpha_k, \beta_k$  and  $\gamma_k$ , this method yields better convergence rate than the standard gradient descent. And in [12], Y. Nesterov obtained the way of choosing the parameter  $\alpha_k, \beta_k$  and  $\gamma_k$ , we set  $\gamma_{-1} = 0$ , choose  $\gamma_k$  to be the larger root of

$$\gamma_k^2 - \frac{\gamma_k}{m} = (1 - \frac{\gamma_k \lambda}{m}) \gamma_{k-1}^2, \quad (8)$$

Then we set  $\alpha_k$  and  $\beta_k$  as follows:

$$\alpha_k = \frac{m - \gamma_k \lambda}{\gamma_k (m^2 - \lambda)}, \beta_k = 1 - \frac{\gamma_k \lambda}{m}. \quad (9)$$

In [10], J. Liu and S. J. Wright showed the ARK method to address the consistent linear system. Since the main computation of REK is about  $4m + 4n$  operations per iteration then the cost per iteration of Algorithm 2 is about  $11n$  with the same assumption as the REK method, incurred in steps of Nesterov's accelerated procedure.

**Remark 1.** If the data matrix A is sparse, with a fraction of  $\delta$  nonzeros (with  $0 < \delta \leq 1$ ), the ARK method has  $s$  significant disadvantage. The average number of operations for each iteration of the RK method is  $O(\delta n)$ , however, since the vectors  $\mathbf{x}_k, \mathbf{y}_k$  and  $\mathbf{v}_k$  are dense in general, the operation of the ARK algorithm remains at  $O(n)$ . Therefore, in this paper, we assume that the data matrix A is dense.

### 2.4 Nesterov's Accelerated Randomized extended Kaczmarz(AREK)

We present Nesterov's accelerated randomized extended Kaczmarz algorithm which is a specific combination of the REK algorithm together with the ARK algorithm. The proposed algorithm consists of two components. The first component is the randomized approximate orthogonal projection which implicitly maintains an approximation to  $\mathbf{b}_{\mathcal{R}(A)}$  formed by  $\mathbf{b} - \mathbf{z}_k$ . The second part applies the ARK method, with Nesterov's accelerated procedure, on the system  $A\mathbf{x} = \mathbf{b} - \mathbf{z}_k$ .

**Table 3.** Nesterov's Accelerated Randomized extended Kaczmarz(AREK) (Algorithm 3)

<b>Require:</b>	$m \times n$ matrix $A$ , coefficient vector $\mathbf{b} \in \mathbb{R}^m$
<b>Initialization:</b>	$\mathbf{v}_0 = \mathbf{x}_0 = \mathbf{0}, \mathbf{z}_0 = \mathbf{b}, \gamma_{-1} = 0.$
<b>Select 1:</b>	Pick $j_k \in [n]$ with probability $p_j := \frac{\ \mathbf{a}^j\ _2^2}{\ A\ _F^2}, j \in [n]$
<b>Project 1:</b>	Set $\mathbf{z}_{k+1} = \mathbf{z}_k - \frac{\langle \mathbf{a}^{j_k}, \mathbf{z}_k \rangle}{\ \mathbf{a}^{j_k}\ _2^2} \mathbf{a}^{j_k}.$ Choose $\gamma_k$ to be the larger root as in (8). Set $\alpha_k$ and $\beta_k$ as in (9).
<b>Select 2:</b>	Pick $i_k \in [m]$ with probability $q_i := \frac{\ \mathbf{a}_i\ _2^2}{\ A\ _F^2}, i \in [m]$
<b>Project 2:</b>	$\mathbf{y}_k = \alpha_k \mathbf{v}_k + (1 - \alpha_k) \mathbf{x}_k; \mathbf{x}_{k+1} = \mathbf{y}_k - \mathbf{a}_{i_k} (\mathbf{a}_{i_k}^T \mathbf{y}_k - (\mathbf{b}^{i_k} - \mathbf{z}_k^{i_k})) / \ \mathbf{a}_{i_k}\ _2^2;$ $\mathbf{v}_{k+1} = \beta_k \mathbf{v}_k + (1 - \beta_k) \mathbf{y}_k - \gamma_k \mathbf{a}_{i_k} (\mathbf{a}_{i_k}^T \mathbf{y}_k - (\mathbf{b}^{i_k} - \mathbf{z}_k^{i_k})) / \ \mathbf{a}_{i_k}\ _2^2.$
<b>Update:</b>	Set $k = k + 1$
<b>end</b>	

The main computation of Algorithm, see Table 3, consists two parts, the first is  $4m$  which is the cost of randomized approximate orthogonal projection, the second component is  $11n$  of that of the ARK method. Next, we consider the expected rate of convergence AREK.

**Theorem 2.** If data matrix  $A$  is dense and  $\lambda_{\min}(A^T A) \rightarrow 0^+$ , after  $k > 1$  iterations, AREK with input  $A$  and  $\mathbf{b}$  computes  $\mathbf{x}$  vector  $\mathbf{x}^*$  such that

$$\mathbb{E} \|\mathbf{x}_k^* - \mathbf{x}^*\|_2^2 \leq \frac{4m^2 \|\mathbf{x}_0 - \mathbf{x}^*\|_{(A^T A)}^2}{k^2} + \left(1 - \frac{1}{R}\right)^k \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 \frac{\sigma_{\max}^2}{\|A\|_F^2}. \quad (10)$$

Before proving Theorem 2, it is first really important to analyze that, when the error vector is added, what will happen to the solution space of the original equations. Next lemma will show that each hyperplane, which is  $H_i = \{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle = \mathbf{b}_i\}$  is shifted, and the distance of each hyperplane.

**Lemma 1.** [6]  $H_i = \{x : \langle a_i, x \rangle = b_i\}$  is the affine subspace of the  $\mathbb{R}^n$  which consist the solution of the unperturbed equations, is the solution space of the noisy equations. Then  $H_i^* = \{\omega + \alpha_i a_i : \omega \in H_i\}$  where  $\alpha_i = \frac{r_i}{\|a_i\|_2^2}$ .

Then we will utilize the convergence rate of the part of the randomized approximate orthogonal projection to obtain convergence rate of our approach.

**Lemma 2.** [8] Let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  be the variable of linear system  $Ax = b$ , after  $k > 0$  iterations of Algorithm 3 it holds that

$$\mathbb{E} \left\| z_k - b_{\mathcal{R}(A)^\perp} \right\|_2^2 \leq \left( 1 - \frac{1}{R} \right)^k \|b_{\mathcal{R}(A)}\|_2^2, \quad (11)$$

Proof of Theorem 2. We denote  $x_k^*$  be the  $k^{\text{th}}$  iterate of the noisy AREK method, and  $x_k$  is the  $k^{\text{th}}$  iterate of the noiseless AREK method. Then we have

$$\mathbb{E} \|x_k^* - x^*\|_2^2 = \mathbb{E} \|x_k^* - x_k + x_k - x^*\|_2^2 = \mathbb{E} \|x_k^* - x_k\|_2^2 + \mathbb{E} \|x_k - x^*\|_2^2.$$

Let  $y_{k-1}^*$  denote the  $(k-1)^{\text{th}}$  iterate of the noisy AREK method. We use the notation as in Lemma 1,  $H_i^*$  and  $H_i$  are the solution spaces chosen in the  $k^{\text{th}}$  iteration. Again by orthogonality, we have

$\mathbb{E} \|x_k^* - x_k\|_2^2 = \mathbb{E} \|\alpha_i a_i\|_2^2$ . If data matrix  $A$  is dense and  $\lambda_{\min}(A^T A) \rightarrow 0^+$ , then by (6), we obtain

$$\mathbb{E} \|x_k^* - x^*\|_2^2 = \mathbb{E} \|x_k - x^*\|_2^2 + \mathbb{E} \|\alpha_i a_i\|_2^2 \leq \frac{4m^2 \|x_0 - x^*\|_{(A^T A)^+}^2}{k^2} + \mathbb{E} \|\alpha_i a_i\|_2^2.$$

Since  $\alpha_i = \frac{r_i}{\|a_i\|_2^2}$  and  $\|a_i\|_2 = 1$ , we have  $\mathbb{E} \|\alpha_i a_i\|_2^2 = \frac{\|r\|_2^2}{\|A\|_F^2}$ .

After  $k$  iterations, we obtain  $b - z_k \approx b_{\mathcal{R}(A)}$ , then we have  $r = z_k - b_{\mathcal{R}(A)^\perp}$ . By Lemma 2,

$$\frac{\|r\|_2^2}{\|A\|_F^2} = \frac{\|b_{\mathcal{R}(A)^\perp} - z_k\|_2^2}{\|A\|_F^2} \leq \left( 1 - \frac{1}{R} \right)^k \frac{\|b_{\mathcal{R}(A)}\|_2^2}{\|A\|_F^2}. \text{ Since } \|b_{\mathcal{R}(A)}\|_2 \leq \sigma_{\max} \|x_0 - x^*\|_2, \text{ we obtain (10).}$$

If data matrix  $A$  is dense and  $\lambda_{\min}(A^T A) \rightarrow 0^+$ , by comparing the conclusion of Theorem 1 and Theorem 2, we conclude that our approach is faster than the REK method.

### 3. Numerical results

In the section, we describe some numerical results for the REK algorithm and the AREK algorithm. We compare REK and AREK for dense  $A$  and  $\lambda_{\min}(A^T A) \rightarrow 0^+$ . Synthetic data for our numerical

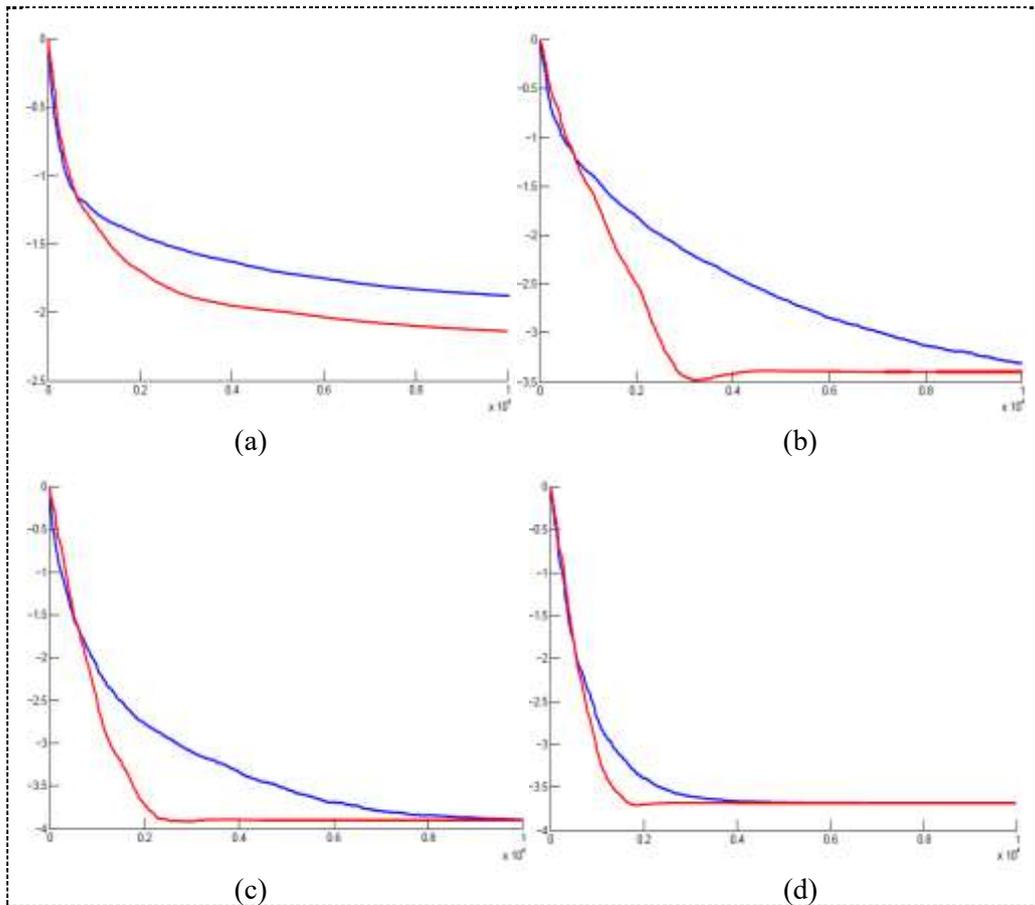
computation is generated as follows: optimal solution  $\mathbf{x}^* \in \mathbb{R}^n$  is chosen to be i.i.d.  $N(0,1)$ . Each row of  $A$  are normalized. We set the right-hand side  $\mathbf{b} = A\mathbf{x}^* + \mathbf{r}$ , and we set  $\mathbf{r}$  i.i.d.  $(1/100)N(0,1)$  and the data matrix  $A \in \mathbb{R}^{m \times n}$  i.i.d.  $(1/10)N(0,1)$ . We measure performance by plotting residual  $\|\mathbf{x}_0 - \mathbf{x}^*\|_2$  against the number of iterations, and the initial point  $\mathbf{x}_0 = 0$  is used in all numerical computation. Figure 1 show residual errors for REK and AREK with different  $\lambda_{\min}(A^T A)$  and  $n$ .

#### 4. Conclusion

This paper proposes an accelerated randomized extended Kaczmarz (AREK) algorithm via Nesterov's accelerated procedure, and we obtain better convergence rate. Both REK and AREK have almost the same complexities per iteration for the dense data. In the section of numerical results, under the assumption of dense data matrix  $A$  and  $\lambda_{\min}(A^T A) \rightarrow 0^+$ , a comparison of convergence rate of REK and AREK is provided.

#### 5. References

- [1] S Kaczmarz 1937 Angenaherte auflösung von system linearer gleichungen *Bulletin International de l'Académie Polonaise des Sciences et des Letter*, vol. 35, pp. 355-357
- [2] H G Feichtinger, C Cenkler, M Mayer, H Steier and T Strohmer 1992 New Variants of the POCS Method using Affine Subspaces of Finite Codimension with Applications to Irregular Sampling *Proceedings of SPIE*, 1818:299-310
- [3] F Natterer 2002 The Mathematics of Computerized Tomography *Medical Physics*, 29(1):107-108
- [4] T Strohmer, R Vershynin 2006 A randomized solver for linear system with exponential Convergence *In RANDOM 2006 (10th International Workshop on Randomization and Computation)*, number 4110 in Lecture Notes in Computer Science, pp. 499-507.
- [5] T Strohmer, R Vershynin 2009 A randomized Kaczmarz algorithm with exponential Convergence *J. Fourier Anal. Appl.*, 15:262-278
- [6] D Needell 2010 Randomized Kaczmarz solver for noisy linear system *BIT Numerical Mathematics*, vol. 50, no. 2, pp. 1422-1436
- [7] Y C Eldar and D Needell 2011 Acceleration of randomized Kaczmarz method via the JohnsonLindenstrauss lemma *Numerical Algorithms*, vol. 58, no. 2, pp. 163-177
- [8] A Zouzias and N M Freris 2012 Randomized extended Kaczmarz for solving least-squares *arXiv:1205.5770v2*
- [9] C Popa 1999 Characterization of the solution set of least-squares problems by an extension of Kaczmarz's projection method *Journal of Applied Mathematics and Computing*, vol. 6, pp. 51-64
- [10] J Liu and S J Wright 2013 An Accelerated Randomized Kaczmarz Algorithm *arXiv:1310.2887v1*
- [11] Y Nesterov 2004 Introductory Lectures on Convex Optimization: A Basic Course *Kluwer Academic Publishers*
- [12] Y Nesterov 2012 Efficiency of coordinate descent methods on huge-scale optimization Problems *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 341-362



**Figure 1.** y-axis denotes  $\log_{10} \|x_0 - x^*\|_2$ , x-axis denotes the number of iterations.

Blue line is REK, red line is AREK. (a):  $m=100$ ,  $n=90$ ,  $\lambda_{\min}(A^T A) = 0.0018$ .

(b):  $m=100$ ,  $n=80$ ,  $\lambda_{\min}(A^T A) = 0.0143$ . (c):  $m=100$ ,  $n=70$ ,  $\lambda_{\min}(A^T A) = 0.0267$ .

(d):  $m=100$ ,  $n=60$ ,  $\lambda_{\min}(A^T A) = 0.0644$ .