SIAM J. SCI. COMPUT. © 2020 Society for Industrial and Applied Mathematics
Vol. 42, No. 6, pp. A3541–A3559

RANDOMIZED EXTENDED AVERAGE BLOCK KACZMARZ FOR SOLVING LEAST SQUARES*

KUI DU[†], WU-TAO SI[‡], AND XIAO-HUI SUN[‡]

Abstract. Randomized iterative algorithms have recently been proposed to solve large-scale linear systems. In this paper, we present a simple randomized extended average block Kaczmarz algorithm that exponentially converges in the mean square to the unique minimum norm least squares solution of a given linear system of equations. The proposed algorithm is pseudoinverse-free and therefore different from the projection-based randomized double block Kaczmarz algorithm of Needell, Zhao, and Zouzias [*Linear Algebra Appl.*, 484 (2015), pp. 322–343]. We emphasize that our method works for all types of linear systems (consistent or inconsistent, overdetermined or underdetermined, full-rank or rank-deficient). Moreover, our approach can be implemented for parallel computation, yielding remarkable improvements in computational time. Numerical examples are given to show the efficiency of the new algorithm.

Key words. general linear systems, minimum norm least squares solution, randomized extended average block Kaczmarz, exponential convergence

AMS subject classifications. 65F10, 65F20

DOI. 10.1137/20M1312629

1. Introduction.

The Kaczmarz method [27] is a simple iterative method for solving a linear system of equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{m\times n}, \quad \mathbf{b} \in \mathbb{R}^m.$$

Due to its simplicity and numerical performance, the Kaczmarz method has found many applications in many fields, such as computer tomography [35, 28, 24], image reconstruction [46, 25], digital signal processing [9, 32], etc. At each step, the method projects the current iterate onto one hyperplane defined by a row of the system. More precisely, assuming that the ith row $\mathbf{A}_{i,:}$ has been selected at the kth iteration, the kth estimate vector $\mathbf{x}^k$ is obtained by

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \alpha_k \frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - \mathbf{b}_i}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^{\mathrm{T}},$$

where $(\mathbf{A}_{i,:})^{\mathrm{T}}$ denotes the transpose of $\mathbf{A}_{i,:}$, $\mathbf{b}_i$ is the ith component of $\mathbf{b}$, $\|\cdot\|_2$ is the Euclidean norm, and α_k is a stepsize. Numerical experiments show that using the rows of the coefficient matrix in the Kaczmarz method in random order rather than in their given order can often greatly improve the convergence [26, 35]. In a seminal paper [49], Strohmer and Vershynin proposed a randomized Kaczmarz (RK) algorithm which exponentially converges in expectation to the solutions of consistent, overdetermined,

*Submitted to the journal's Methods and Algorithms for Scientific Computing section January 13, 2020; accepted for publication (in revised form) August 7, 2020; published electronically November 2, 2020.
https://doi.org/10.1137/20M1312629
Funding: The research of the first author was supported by the National Natural Science Foundation of China (11771364) and the Fundamental Research Funds for the Central Universities (20720180008).

[†]School of Mathematical Sciences and Fujian Provincial Key Laboratory of Mathematical Modelling and High Performance Scientific Computing, Xiamen University, Xiamen 361005, China (kuidu@xmu.edu.cn).
[‡]School of Mathematical Sciences, Xiamen University, Xiamen 361005, China (siwutao@stu.xmu.edu.cn, sunxh@stu.xmu.edu.cn).

full-rank linear systems. The convergence result was extended and refined in various directions including inconsistent [29, 38, 52, 17, 42, 39, 21], underdetermined or rank-deficient linear systems [33, 20, 47, 15], ridge regression problems [23, 31], linear feasibility problems [11], convex feasibility problems [37], block variants [40, 41, 36], acceleration strategies [30, 48, 3, 4, 5, 6, 51], and many others [2, 45, 13, 14, 22].

Let $\mathbf{A}^\dagger$ denote the Moore–Penrose pseudoinverse[1] [7] of $\mathbf{A}$. In this paper, we are interested in the vector $\mathbf{A}^\dagger\mathbf{b}$. Here we would like to make clear what $\mathbf{A}^\dagger\mathbf{b}$ stands for different types of linear systems (see [7, 19]):

(1) If $\mathbf{Ax}=\mathbf{b}$ is consistent with full-column rank $\mathbf{A}$, i.e., $\operatorname{rank}(\mathbf{A})=n$, then $\mathbf{A}^\dagger\mathbf{b}$ is the unique solution. In this case, we have $m\geq n$, and the linear system is overdetermined when $m>n$.

(2) If $\mathbf{Ax}=\mathbf{b}$ is consistent with $\operatorname{rank}(\mathbf{A})<n$, then $\mathbf{A}^\dagger\mathbf{b}$ is the unique minimum norm solution. In this case, we have $m\geq n$ or $m<n$, and the linear system is overdetermined (resp., underdetermined) when $m>n$ (resp., $m<n$). The matrix $\mathbf{A}$ can be of full-row rank, i.e., $\operatorname{rank}(\mathbf{A})=m$, or rank-deficient, i.e., $\operatorname{rank}(\mathbf{A})<m$.

(3) If $\mathbf{Ax}=\mathbf{b}$ is inconsistent with $\operatorname{rank}(\mathbf{A})=n$, then $\mathbf{A}^\dagger\mathbf{b}$ is the unique least squares solution. In this case, we have $m\geq n$, and the linear system is overdetermined when $m>n$.

(4) If $\mathbf{Ax}=\mathbf{b}$ is inconsistent with $\operatorname{rank}(\mathbf{A})<n$, then $\mathbf{A}^\dagger\mathbf{b}$ is the unique minimum norm least squares solution. In this case, we have $m\geq n$ or $m<n$, and the linear system is overdetermined (resp., underdetermined) when $m>n$ (resp., $m<n$). The matrix $\mathbf{A}$ can be of full-row rank, i.e., $\operatorname{rank}(\mathbf{A})=m$, or rank-deficient, i.e., $\operatorname{rank}(\mathbf{A})<m$.

If $\mathbf{Ax}=\mathbf{b}$ is inconsistent, Needell [38] showed that RK does not converge to $\mathbf{A}^\dagger\mathbf{b}$. To resolve this problem, Zouzias and Freris [52] proposed a randomized extended Kaczmarz (REK) algorithm which uses RK twice [30, 13] at each iteration and exponentially converges in the mean square to $\mathbf{A}^\dagger\mathbf{b}$. More precisely, assuming that the jth column $\mathbf{A}_{:,j}$ and the ith row $\mathbf{A}_{i,:}$ have been selected at the kth iteration, REK generates two vectors $\mathbf{z}^k$ and $\mathbf{x}^k$ via two RK updates (one for $\mathbf{A}^{\mathrm{T}}\mathbf{z}=\mathbf{0}$ from $\mathbf{z}^{k-1}$ and the other for $\mathbf{Ax}=\mathbf{b}-\mathbf{z}^k$ from $\mathbf{x}^{k-1}$):

$$\mathbf{z}^k=\mathbf{z}^{k-1}-\frac{(\mathbf{A}_{:,j})^{\mathrm{T}}\mathbf{z}^{k-1}}{\|\mathbf{A}_{:,j}\|_2^2}\mathbf{A}_{:,j},$$
$$\mathbf{x}^k=\mathbf{x}^{k-1}-\frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1}-\mathbf{b}_i+\mathbf{z}_i^k}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^{\mathrm{T}}.$$

For general linear systems (consistent or inconsistent, full-rank or rank-deficient), the vector $\mathbf{x}^k$ generated by REK exponentially converges to $\mathbf{A}^\dagger\mathbf{b}$ if $\mathbf{z}^0\in\mathbf{b}+\operatorname{range}(\mathbf{A})$ and $\mathbf{x}^0\in\operatorname{range}(\mathbf{A}^{\mathrm{T}})$ [30, 13]. To accelerate the convergence, the following projection-based block variants [40, 41] of RK and REK were developed. For a subset $\mathcal{I}\subset\{1,2,\ldots,m\}$ and a subset $\mathcal{J}\subset\{1,2,\ldots,n\}$, denote by $\mathbf{A}_{\mathcal{I},:}$ and $\mathbf{A}_{:,\mathcal{J}}$ the row submatrix of $\mathbf{A}$ indexed by $\mathcal{I}$ and the column submatrix of $\mathbf{A}$ indexed by $\mathcal{J}$, respectively. Assuming that the subset $\mathcal{I}_i$ has been selected at the kth iteration, the randomized block Kaczmarz (RBK) algorithm [40] generates the kth estimate $\mathbf{x}^k$ via

$$\mathbf{x}^k=\mathbf{x}^{k-1}-(\mathbf{A}_{\mathcal{I}_i,:})^\dagger(\mathbf{A}_{\mathcal{I}_i,:}\mathbf{x}^{k-1}-\mathbf{b}_{\mathcal{I}_i}).$$

[1]Every $m\times n$ matrix $\mathbf{A}$ has a unique Moore–Penrose pseudoinverse. In particular, in this paper we will use the following property of the pseudoinverse: $\mathbf{A}^{\mathrm{T}}=\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{A}^\dagger$.

Assuming that the subsets $\mathcal{J}_j$ and $\mathcal{I}_i$ have been selected at the k iteration, the randomized double block Kaczmarz (RDBK) algorithm [41] generates the kth estimate $\mathbf{x}^k$ via

$$\begin{aligned}\mathbf{z}^k &= \mathbf{z}^{k-1} - \mathbf{A}_{:,\mathcal{J}_j}(\mathbf{A}_{:,\mathcal{J}_j})^\dagger \mathbf{z}^{k-1},\\ \mathbf{x}^k &= \mathbf{x}^{k-1} - (\mathbf{A}_{\mathcal{I}_i,:})^\dagger(\mathbf{A}_{\mathcal{I}_i,:}\mathbf{x}^{k-1} - \mathbf{b}_{\mathcal{I}_i} + \mathbf{z}^k_{\mathcal{I}_i}).\end{aligned}$$

Numerical experiments demonstrate that the convergence can be significantly accelerated if appropriate blocks of the coefficient matrix are used. The main drawback of projection-based block methods is that they are difficult to parallelize.

Recently, Necoara [36] proposed a randomized average block Kaczmarz (RABK) algorithm for consistent linear systems which takes a convex combination of several RK updates (i.e., the projections of the current iterate onto several hyperplanes) as a new direction with some stepsize. Assuming that the subset $\mathcal{I}$ has been selected at the kth iteration, RABK generates the kth estimate $\mathbf{x}^k$ via

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \alpha_k \left(\sum_{i\in\mathcal{I}} \omega_i^k \frac{\mathbf{A}_{i,:}\mathbf{x}^{k-1} - \mathbf{b}_i}{\|\mathbf{A}_{i,:}\|_2^2} (\mathbf{A}_{i,:})^{\mathrm{T}} \right), \tag{1.1}$$

where the weights $\omega_i^k \in [0,1]$ such that $\sum_{i\in\mathcal{I}} \omega_i^k = 1$ and the stepsize $\alpha_k \in (0,2)$. The convergence analysis reveals that RABK is extremely effective when it is given a good sampling of the rows into well-conditioned blocks. Moorman et al. [34] developed a theory for RABK (which is referred to as "randomized Kaczmarz with averaging" in their paper) when applied to inconsistent systems. A block version of RABK (i.e., parallel RBK) which takes a convex combination of the RBK updates was proposed and studied by Richtárik and Takáč [48]. Shortly afterwards, Du and Sun [15] proposed a doubly stochastic block Gauss–Seidel (DSBGS) algorithm which randomly chooses a submatrix of the coefficient matrix at each iteration. Assuming that the subsets $\mathcal{I}$ and $\mathcal{J}$ have been selected at the kth iteration, DSBGS generates the kth estimate $\mathbf{x}^k$ via

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \alpha_k \frac{\mathbf{I}_{:,\mathcal{J}}(\mathbf{A}_{\mathcal{I},\mathcal{J}})^{\mathrm{T}}(\mathbf{I}_{:,\mathcal{I}})^{\mathrm{T}}}{\|\mathbf{A}_{\mathcal{I},\mathcal{J}}\|_{\mathrm{F}}^2}(\mathbf{A}\mathbf{x}^{k-1} - \mathbf{b}),$$

where $\mathbf{I}$ denotes the identity matrix, $\mathbf{A}_{\mathcal{I},\mathcal{J}}$ denotes the submatrix that lies in the rows indexed by $\mathcal{I}$ and the columns indexed by $\mathcal{J}$, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. Exponential convergence of DSBGS for consistent linear systems was proved. By setting $\mathcal{I} \subset \{1,2,\ldots,m\}$ and $\mathcal{J} = \{1,2,\ldots,n\}$, DSBGS recovers a special case of RABK, i.e., RABK with weight

$$\omega_i^k = \frac{\|\mathbf{A}_{i,:}\|_2^2}{\|\mathbf{A}_{\mathcal{I},:}\|_{\mathrm{F}}^2}, \qquad i \in \mathcal{I}.$$

A feature of the RABK algorithm is that it allows one to project in parallel onto several rows, yielding remarkable improvements in computational time. We note that the existing convergence results show that RABK applied to inconsistent linear systems converges only to within a radius (*convergence horizon*) of the least squares solution [34].

In this paper, based on the REK algorithm and the RABK algorithm, we present a simple randomized extended average block Kaczmarz (REABK) algorithm that exponentially converges in the mean square to the unique minimum norm (least squares) solution of a given general linear system (full-rank or rank-deficient, overdetermined or underdetermined, consistent or inconsistent). Our method is different from those projection-based block methods, for example, those in [18, 1, 8, 44, 40, 41, 16]. At each step, REABK, as a direct extension of REK, uses two special RABK (which also can be viewed as special DSBGS) updates (one for $\mathbf{A}^{\mathrm{T}}\mathbf{z} = \mathbf{0}$ from $\mathbf{z}^{k-1}$ and the other for $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^k$ from $\mathbf{x}^{k-1}$; see section 2 for details). Compared with REK, REABK usually has a better convergence rate and can exploit the high-level basic linear algebra subroutine (BLAS2), even fast matrix-vector multiplies (for example, if submatrices of $\mathbf{A}$ have circulant or Toeplitz structures, then the fast Fourier transform technique can be used), and therefore could be more efficient. Compared with RDBK, REABK can be implemented for parallel computation. We refer the reader to [40, 36] for more advantages of block methods. In this paper, we give numerical examples to illustrate the efficiency of REABK.

Organization of the paper. In the rest of this section, we give some notation. In section 2 we describe the REABK algorithm and prove its convergence theory. The exponential convergence of both the norm of the expected error and the expected squared norm of the error is discussed. In section 3 we report the numerical results. Finally, we present brief concluding remarks in section 4.

Notation. For any random variable $\boldsymbol{\xi}$, let $\mathbb{E}\left[\boldsymbol{\xi}\right]$ denote its expectation. For an integer $m \geq 1$, let $[m] := \{1, 2, 3, \dots, m\}$. Lowercase (uppercase) boldface letters are reserved for column vectors (matrices). For any vector $\mathbf{u} \in \mathbb{R}^m$, we use $\mathbf{u}_i$, $\mathbf{u}^{\mathrm{T}}$, and $\|\mathbf{u}\|_2$ to denote the ith element, the transpose, and the Euclidean norm of $\mathbf{u}$, respectively. We use $\mathbf{I}$ to denote the identity matrix whose order is clear from the context. For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, we use $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\dagger}$, $\|\mathbf{A}\|_{\mathrm{F}}$, range$(\mathbf{A})$, rank$(\mathbf{A})$, $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) > 0$ to denote the transpose, the Moore–Penrose pseudoinverse, the Frobenius norm, the column space, the rank, and all the nonzero singular values of $\mathbf{A}$, respectively. We also denote the largest and smallest nonzero singular values of $\mathbf{A}$ by $\sigma_{\max}(\mathbf{A})$ and $\sigma_{\min}(\mathbf{A})$, respectively. For index sets $\mathcal{I} \subseteq [m]$ and $\mathcal{J} \subseteq [n]$, let $\mathbf{A}_{\mathcal{I},:}$, $\mathbf{A}_{:,\mathcal{J}}$, and $\mathbf{A}_{\mathcal{I},\mathcal{J}}$ denote the row submatrix indexed by $\mathcal{I}$, the column submatrix indexed by $\mathcal{J}$, and the submatrix that lies in the rows indexed by $\mathcal{I}$ and the columns indexed by $\mathcal{J}$, respectively. We call $\{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_s\}$ a partition of $[m]$ if $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{s}\mathcal{I}_i = [m]$. Similarly, $\{\mathcal{J}_1, \mathcal{J}_2, \dots, \mathcal{J}_t\}$ denotes a partition of $[n]$ if $\mathcal{J}_i \cap \mathcal{J}_j = \emptyset$ for $i \neq j$ and $\cup_{j=1}^{t}\mathcal{J}_j = [n]$. We use $|\mathcal{I}|$ to denote the cardinality of a set $\mathcal{I} \subseteq [m]$.

2. The REABK algorithm.

In this section, based on given partitions of $[m]$ and $[n]$, we propose the following REABK algorithm (see Algorithm 1) for solving consistent or inconsistent linear systems. We emphasize that this algorithm can be implemented for parallel computation.

Here we only consider constant stepsize for simplicity. By choosing the row partition parameter $s = m$, the column partition parameter $t = n$, and the stepsize $\alpha = 1$, we recover the well-known REK algorithm of Zouzias and Freris [52]. REABK uses two RABK updates (see (1.1)) at each step:

- RABK update for $\mathbf{A}^{\mathrm{T}}\mathbf{z} = \mathbf{0}$ from $\mathbf{z}^{k-1}$

$$\mathbf{z}^k = \mathbf{z}^{k-1} - \alpha \left(\sum_{l\in\mathcal{J}_j} \omega_l \frac{(\mathbf{A}_{:,l})^{\mathrm{T}}\mathbf{z}^{k-1}}{\|\mathbf{A}_{:,l}\|_2^2} \mathbf{A}_{:,l} \right), \qquad \omega_l = \frac{\|\mathbf{A}_{:,l}\|_2^2}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2};$$

- RABK update for $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^k$ from $\mathbf{x}^{k-1}$

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \alpha \left(\sum_{l \in \mathcal{I}_i} \omega_l \frac{\mathbf{A}_{l,:}\mathbf{x}^{k-1} - \mathbf{b}_l + \mathbf{z}_l^k}{\|\mathbf{A}_{l,:}\|_2^2} (\mathbf{A}_{l,:})^{\mathrm{T}} \right), \qquad \omega_l = \frac{\|\mathbf{A}_{l,:}\|_2^2}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}.$$

We note that if $\mathbf{z}^0 = \mathbf{0}$ in REABK, then all $\mathbf{z}^k \equiv \mathbf{0}$. This yields the update $\mathbf{x}^k$ that is exactly the same as that of RABK.

Before proving the convergence theory of REABK for general linear systems, we give the following notation. Let $\mathbb{E}_{k-1}\left[\cdot\right]$ denote the conditional expectation conditioned on the first $k-1$ iterations of REABK. That is,

$$\mathbb{E}_{k-1}\left[\cdot\right] = \mathbb{E}\left[\cdot | j_1, i_1, j_2, i_2, \ldots, j_{k-1}, i_{k-1}\right],$$

where j_l is the lth column block chosen and i_l is the lth row block chosen. We denote the conditional expectation conditioned on the first $k-1$ iterations and the kth column block chosen as

$$\mathbb{E}_{k-1}^i\left[\cdot\right] = \mathbb{E}\left[\cdot | j_1, i_1, j_2, i_2, \ldots, j_{k-1}, i_{k-1}, j_k\right].$$

Then by the law of total expectation we have

$$\mathbb{E}_{k-1}\left[\cdot\right] = \mathbb{E}_{k-1}\left[\mathbb{E}_{k-1}^i\left[\cdot\right]\right].$$

Algorithm 1: Randomized extended average block Kaczmarz (REABK)

Let $\{\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_s\}$ and $\{\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_t\}$ be partitions of $[m]$ and $[n]$, respectively.
Let $\alpha > 0$. Initialize $\mathbf{z}^0 \in \mathbb{R}^m$ and $\mathbf{x}^0 \in \mathbb{R}^n$.
for $k = 1, 2, \ldots,$ **do**
  Pick $j \in [t]$ with probability $\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2 / \|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{z}^k = \mathbf{z}^{k-1} - \dfrac{\alpha}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2} \mathbf{A}_{:,\mathcal{J}_j} (\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}} \mathbf{z}^{k-1}$
  Pick $i \in [s]$ with probability $\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2 / \|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \dfrac{\alpha}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2} (\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}} (\mathbf{A}_{\mathcal{I}_i,:}\mathbf{x}^{k-1} - \mathbf{b}_{\mathcal{I}_i} + \mathbf{z}_{\mathcal{I}_i}^k)$

2.1. The exponential convergence of the norm of the expected error. In this subsection we show the exponential convergence of the norm of the expected error, i.e.,

$$\|\mathbb{E}\left[\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b}\right]\|_2.$$

The convergence of the norm of the expected error depends on the nonnegative number δ defined as

$$\delta := \max_{1 \le i \le r} \left| 1 - \frac{\alpha \sigma_i^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2} \right|.$$

The following lemma will be used, and its proof is straightforward (e.g., via the singular value decomposition).

LEMMA 2.1. *Let* $\alpha > 0$ *and* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be any nonzero real matrix with* $\operatorname{rank}(\mathbf{A}) = r$. *For every* $\mathbf{u} \in \operatorname{range}(\mathbf{A}^{\mathrm{T}})$, *it holds that*

$$\left\| \left(\mathbf{I} - \alpha \frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2} \right)^k \mathbf{u} \right\|_2 \le \delta^k \|\mathbf{u}\|_2.$$

We give the convergence of the norm of the expected error of REABK in the following theorem.

THEOREM 2.2. *For any given consistent or inconsistent linear system* $\mathbf{Ax} = \mathbf{b}$, *let* $\mathbf{x}^k$ *be the kth iterate of* REABK *with* $\mathbf{z}^0 \in \mathbb{R}^m$ *and* $\mathbf{x}^0 \in \operatorname{range}(\mathbf{A}^{\mathrm{T}})$. *It holds that*

$$\|\mathbb{E}\left[\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\right]\|_2 \leq \delta^k \left(\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2 + \frac{\alpha k \|\mathbf{A}^{\mathrm{T}}\mathbf{z}^0\|_2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right).$$

Proof. Note that

$$\begin{aligned}\mathbb{E}_{k-1}\left[\mathbf{z}^k\right] &= \mathbf{z}^{k-1} - \mathbb{E}_{k-1}\left[\frac{\alpha}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2}\mathbf{A}_{:,\mathcal{J}_j}(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}\right]\mathbf{z}^{k-1} \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbf{z}^{k-1},\end{aligned}$$

and therefore

$$\mathbb{E}\left[\mathbf{z}^k\right] = \mathbb{E}\left[\mathbb{E}_{k-1}\left[\mathbf{z}^k\right]\right] = \left(\mathbf{I} - \alpha\frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbb{E}\left[\mathbf{z}^{k-1}\right] = \left(\mathbf{I} - \alpha\frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k \mathbf{z}^0.$$

By $\mathbf{A}^{\mathrm{T}}\mathbf{b} = \mathbf{A}^{\mathrm{T}}\mathbf{AA}^\dagger\mathbf{b}$, we have

$$\begin{aligned}&\mathbb{E}_{k-1}\left[\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\right] = \mathbb{E}_{k-1}\left[\mathbb{E}_{k-1}^i\left[\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\right]\right] \\ &= \mathbb{E}_{k-1}\left[\mathbb{E}_{k-1}^i\left[\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\alpha}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}(\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}}(\mathbf{A}_{\mathcal{I}_i,:}\mathbf{x}^{k-1} - \mathbf{b}_{\mathcal{I}_i} + \mathbf{z}_{\mathcal{I}_i}^k)\right]\right] \\ &= \mathbb{E}_{k-1}\left[\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \alpha\frac{\mathbf{A}^{\mathrm{T}}(\mathbf{Ax}^{k-1} - \mathbf{b} + \mathbf{z}^k)}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right] \\ &= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{Ax}^{k-1} - \mathbf{A}^{\mathrm{T}}\mathbf{b}}{\|\mathbf{A}\|_{\mathrm{F}}^2} - \alpha\frac{\mathbf{A}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}_{k-1}\left[\mathbf{z}^k\right] \\ &= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{Ax}^{k-1} - \mathbf{A}^{\mathrm{T}}\mathbf{AA}^\dagger\mathbf{b}}{\|\mathbf{A}\|_{\mathrm{F}}^2} - \alpha\frac{\mathbf{A}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}_{k-1}\left[\mathbf{z}^k\right] \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)(\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}) - \alpha\frac{\mathbf{A}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\left(\mathbf{I} - \alpha\frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbf{z}^{k-1}.\end{aligned}$$

Taking expectation gives

$$\begin{aligned}\mathbb{E}\left[\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\right] &= \mathbb{E}\left[\mathbb{E}_{k-1}\left[\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\right]\right] \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbb{E}\left[\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}\right] - \alpha\frac{\mathbf{A}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\left(\mathbf{I} - \alpha\frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbb{E}\left[\mathbf{z}^{k-1}\right] \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbb{E}\left[\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}\right] - \alpha\frac{\mathbf{A}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\left(\mathbf{I} - \alpha\frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k\mathbf{z}^0 \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbb{E}\left[\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}\right] - \alpha\left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k\frac{\mathbf{A}^{\mathrm{T}}\mathbf{z}^0}{\|\mathbf{A}\|_{\mathrm{F}}^2} \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^2\mathbb{E}\left[\mathbf{x}^{k-2} - \mathbf{A}^\dagger\mathbf{b}\right] - 2\alpha\left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k\frac{\mathbf{A}^{\mathrm{T}}\mathbf{z}^0}{\|\mathbf{A}\|_{\mathrm{F}}^2} \\ &= \cdots \\ &= \left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k(\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}) - \alpha k\left(\mathbf{I} - \alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k\frac{\mathbf{A}^{\mathrm{T}}\mathbf{z}^0}{\|\mathbf{A}\|_{\mathrm{F}}^2}.\end{aligned}$$

Applying the norms to both sides we obtain

$$\begin{aligned}\|\mathbb{E}\left[\mathbf{x}^k-\mathbf{A}^\dagger\mathbf{b}\right]\|_2 &= \left\|\left(\mathbf{I}-\alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k(\mathbf{x}^0-\mathbf{A}^\dagger\mathbf{b})-\alpha k\left(\mathbf{I}-\alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k\frac{\mathbf{A}^{\mathrm{T}}\mathbf{z}^0}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right\|_2\\ &\le \left\|\left(\mathbf{I}-\alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k(\mathbf{x}^0-\mathbf{A}^\dagger\mathbf{b})\right\|_2+\left\|\alpha k\left(\mathbf{I}-\alpha\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)^k\frac{\mathbf{A}^{\mathrm{T}}\mathbf{z}^0}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right\|_2\\ &\le \delta^k\left(\|\mathbf{x}^0-\mathbf{A}^\dagger\mathbf{b}\|_2+\frac{\alpha k\|\mathbf{A}^{\mathrm{T}}\mathbf{z}^0\|_2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right).\end{aligned}$$

Here the last inequality follows from the fact that $\mathbf{x}^0\in\mathrm{range}(\mathbf{A}^{\mathrm{T}})$, $\mathbf{A}^\dagger\mathbf{b}\in\mathrm{range}(\mathbf{A}^{\mathrm{T}})$, and $\mathbf{A}^{\mathrm{T}}\mathbf{z}^0\in\mathrm{range}(\mathbf{A}^{\mathrm{T}})$ and from Lemma 2.1. ∎

Remark 2.3. To ensure convergence of the expected error, it suffices to have

$$\delta=\max_{1\le i\le r}\left|1-\frac{\alpha\sigma_i^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right|<1,$$

which implies

$$0<\alpha<\frac{2\|\mathbf{A}\|_{\mathrm{F}}^2}{\sigma_{\max}^2(\mathbf{A})}.$$

The optimal α in Theorem 2.2 is (see [43])

$$\frac{2\|\mathbf{A}\|_{\mathrm{F}}^2}{\sigma_{\max}^2(\mathbf{A})+\sigma_{\min}^2(\mathbf{A})}=\underset{0<\alpha<\frac{2\|\mathbf{A}\|_{\mathrm{F}}^2}{\sigma_{\max}^2(\mathbf{A})}}{\mathrm{argmin}}\ \max_{1\le i\le r}\left|1-\frac{\alpha\sigma_i^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right|,$$

and the corresponding convergence rate δ is

$$\frac{\sigma_{\max}^2(\mathbf{A})-\sigma_{\min}^2(\mathbf{A})}{\sigma_{\max}^2(\mathbf{A})+\sigma_{\min}^2(\mathbf{A})}.$$

2.2. The exponential convergence of the expected norm of the error. In this subsection we show the exponential convergence of the expected norm of the error, i.e.,

$$\mathbb{E}\left[\|\mathbf{x}^k-\mathbf{A}^\dagger\mathbf{b}\|_2^2\right].$$

The convergence of the expected norm of the error depends on the real numbers η and ρ defined as

$$\eta:=1-\frac{(2\alpha-\alpha^2\beta_{\max}^{\mathcal{I}})\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2},\qquad \rho:=1-\frac{(2\alpha-\alpha^2\beta_{\max}^{\mathcal{J}})\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2},$$

where

$$\beta_{\max}^{\mathcal{I}}:=\max_{i\in[s]}\frac{\sigma_{\max}^2(\mathbf{A}_{\mathcal{I}_i,:})}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2},\qquad \beta_{\max}^{\mathcal{J}}:=\max_{j\in[t]}\frac{\sigma_{\max}^2(\mathbf{A}_{:,\mathcal{J}_j})}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2}.$$

The following lemmas will be used extensively in this paper.

LEMMA 2.4. *Let* $\mathbf{A}\in\mathbb{R}^{m\times n}$ *be any nonzero real matrix. For every* $\mathbf{u}\in\mathrm{range}(\mathbf{A})$, *it holds that*

$$\|\mathbf{A}^{\mathrm{T}}\mathbf{u}\|_2^2\ge\sigma_{\min}^2(\mathbf{A})\|\mathbf{u}\|_2^2.$$

LEMMA 2.5. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *be any nonzero real matrix. For every* $\mathbf{u} \in \mathbb{R}^m$*, it holds that*

$$\|\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{u}\|_2^2 \le \sigma_{\max}^2(\mathbf{A})\|\mathbf{A}^{\mathrm{T}}\mathbf{u}\|_2^2.$$

The proofs of Lemmas 2.4 and 2.5 are straightforward (e.g., via the singular value decomposition). In the following lemma we show that the vector $\mathbf{z}^k$ generated in REABK with $\mathbf{z}^0 \in \mathbf{b} + \mathrm{range}(\mathbf{A})$ converges to

$$\mathbf{b}_\perp =: (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b},$$

which is the orthogonal projection of $\mathbf{z}^0$ onto the set $\{\mathbf{z} \mid \mathbf{A}^{\mathrm{T}}\mathbf{z} = \mathbf{0}\}$.

LEMMA 2.6. *For any given consistent or inconsistent linear system* $\mathbf{A}\mathbf{x} = \mathbf{b}$*, let* $\mathbf{z}^k$ *be the vector generated in* REABK *with* $\mathbf{z}^0 \in \mathbf{b} + \mathrm{range}(\mathbf{A})$*. Assume* $0 < \alpha < 2/\beta_{\max}^{\mathcal{J}}$*. It holds that*

$$\mathbb{E}\left[\|\mathbf{z}^k - \mathbf{b}_\perp\|_2^2\right] \le \rho^k\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2.$$

Proof. By $(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}\mathbf{b}_\perp = \mathbf{0}$, we have

$$\mathbf{z}^k - \mathbf{b}_\perp = \mathbf{z}^{k-1} - \mathbf{b}_\perp - \frac{\alpha}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2}\mathbf{A}_{:,\mathcal{J}_j}(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp). \tag{2.1}$$

By $\mathbf{z}^0 - \mathbf{b}_\perp = \mathbf{A}\mathbf{A}^\dagger\mathbf{z}^0 \in \mathrm{range}(\mathbf{A})$ and $\mathbf{A}_{:,\mathcal{J}_j}(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp) \in \mathrm{range}(\mathbf{A})$, we can show that $\mathbf{z}^k - \mathbf{b}_\perp \in \mathrm{range}(\mathbf{A})$ by induction. It follows from (2.1) that

$$\begin{aligned}
\|\mathbf{z}^k - \mathbf{b}_\perp\|_2^2 &= \|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2 - \frac{2\alpha\|(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp)\|_2^2}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2} \\
&\quad + \alpha^2\left\|\frac{\mathbf{A}_{:,\mathcal{J}_j}}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}}\left(\frac{\mathbf{A}_{:,\mathcal{J}_j}}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}}\right)^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp)\right\|_2^2 \\
&\le \|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2 - \left(2\alpha - \frac{\alpha^2\sigma_{\max}^2(\mathbf{A}_{:,\mathcal{J}_j})}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2}\right)\frac{\|(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp)\|_2^2}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2} \\
&\qquad \text{(by Lemma 2.5)} \\
&\le \|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2 - (2\alpha - \alpha^2\beta_{\max}^{\mathcal{J}})\frac{\|(\mathbf{A}_{:,\mathcal{J}_j})^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp)\|_2^2}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2}.
\end{aligned}$$

Taking the conditioned expectation on the first $k-1$ iterations yields

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\|\mathbf{z}^k - \mathbf{b}_\perp\|_2^2\right] &\le \|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2 - \frac{(2\alpha - \alpha^2\beta_{\max}^{\mathcal{J}})\|\mathbf{A}^{\mathrm{T}}(\mathbf{z}^{k-1} - \mathbf{b}_\perp)\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} \\
&\le \|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2 - \frac{(2\alpha - \alpha^2\beta_{\max}^{\mathcal{J}})\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2 \\
&\qquad \text{(by Lemma 2.4 and } 0 < \alpha < 2/\beta_{\max}^{\mathcal{J}}) \\
&= \rho\|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2.
\end{aligned}$$

Taking expectation again gives

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{z}^k - \mathbf{b}_\perp\|_2^2\right] &= \mathbb{E}\left[\mathbb{E}_{k-1}\left[\|\mathbf{z}^k - \mathbf{b}_\perp\|_2^2\right]\right] \\
&\le \rho\mathbb{E}\left[\|\mathbf{z}^{k-1} - \mathbf{b}_\perp\|_2^2\right] \\
&\le \rho^k\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2.
\end{aligned}$$

This completes the proof. ∎

We give the main convergence result of REABK in the following theorem.

THEOREM 2.7. *For any given consistent or inconsistent linear system* $\mathbf{A}\mathbf{x} = \mathbf{b}$*, let* $\mathbf{x}^k$ *be the kth iterate of* REABK *with* $\mathbf{z}^0 \in \mathbf{b} + \mathrm{range}(\mathbf{A})$ *and* $\mathbf{x}^0 \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$*. Assume that* $0 < \alpha < 2/\max(\beta^{\mathcal{I}}_{\max}, \beta^{\mathcal{J}}_{\max})$*. For any* $\varepsilon > 0$*, it holds that*

$$\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \le (1+\varepsilon)^k\eta^k\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2 + \left(1+\frac{1}{\varepsilon}\right)\frac{\alpha^2\beta^{\mathcal{I}}_{\max}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\sum_{l=0}^{k-1}\rho^{k-l}(1+\varepsilon)^l\eta^l.$$

Proof. Let

$$\widehat{\mathbf{x}}^k = \mathbf{x}^{k-1} - \frac{\alpha}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}(\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}}\mathbf{A}_{\mathcal{I}_i,:}(\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}),$$

which is actually one RABK update for the linear system $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^\dagger\mathbf{b}$ from $\mathbf{x}^{k-1}$. It follows from

$$\mathbf{x}^k - \widehat{\mathbf{x}}^k = \frac{\alpha}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}(\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}}(\mathbf{b}_{\mathcal{I}_i} - \mathbf{A}_{\mathcal{I}_i,:}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k_{\mathcal{I}_i})$$

that

$$\begin{aligned}\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2 &= \frac{\alpha^2}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^4}\|(\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}}(\mathbf{b}_{\mathcal{I}_i} - \mathbf{A}_{\mathcal{I}_i,:}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k_{\mathcal{I}_i})\|_2^2\\ &\le \frac{\alpha^2}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}\frac{\sigma^2_{\max}(\mathbf{A}_{\mathcal{I}_i,:})}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}\|\mathbf{b}_{\mathcal{I}_i} - \mathbf{A}_{\mathcal{I}_i,:}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k_{\mathcal{I}_i}\|_2^2\\ &\le \frac{\alpha^2\beta^{\mathcal{I}}_{\max}}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}\|\mathbf{b}_{\mathcal{I}_i} - \mathbf{A}_{\mathcal{I}_i,:}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k_{\mathcal{I}_i}\|_2^2. \qquad (2.2)\end{aligned}$$

It follows from

$$\begin{aligned}\mathbb{E}_{k-1}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right] &= \mathbb{E}_{k-1}\left[\mathbb{E}^i_{k-1}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right]\right]\\ &\le \mathbb{E}_{k-1}\left[\mathbb{E}^i_{k-1}\left[\frac{\alpha^2\beta^{\mathcal{I}}_{\max}}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}\|\mathbf{b}_{\mathcal{I}_i} - \mathbf{A}_{\mathcal{I}_i,:}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k_{\mathcal{I}_i}\|_2^2\right]\right]\\ &\qquad \text{(by (2.2))}\\ &= \mathbb{E}_{k-1}\left[\frac{\alpha^2\beta^{\mathcal{I}}_{\max}\|\mathbf{b} - \mathbf{A}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right]\end{aligned}$$

that

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right] &= \mathbb{E}\left[\mathbb{E}_{k-1}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right]\right]\\ &\le \frac{\alpha^2\beta^{\mathcal{I}}_{\max}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}\left[\|\mathbf{b} - \mathbf{A}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}^k\|_2^2\right]\\ &\le \frac{\alpha^2\beta^{\mathcal{I}}_{\max}\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2 \quad \text{(by Lemma 2.6)}. \qquad (2.3)\end{aligned}$$

By $\mathbf{x}^0 \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$, $\mathbf{A}^\dagger\mathbf{b} \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$, $(\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}}(\mathbf{A}_{\mathcal{I}_i,:}\mathbf{x}^{k-1} - \mathbf{b}_{\mathcal{I}_i} + \mathbf{z}^k_{\mathcal{I}_i}) \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$, and

$$\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b} = \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\alpha}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2}(\mathbf{A}_{\mathcal{I}_i,:})^{\mathrm{T}}(\mathbf{A}_{\mathcal{I}_i,:}\mathbf{x}^{k-1} - \mathbf{b}_{\mathcal{I}_i} + \mathbf{z}^k_{\mathcal{I}_i}),$$

we can show that $\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b} \in \text{range}(\mathbf{A}^{\mathrm{T}})$ by induction. By

$$\begin{aligned}
\|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2 &= \|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2 - \frac{2\alpha \|\mathbf{A}_{\mathcal{I}_i,:}(\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b})\|_2^2}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2} \\
&\quad + \alpha^2 \left\| \left(\frac{\mathbf{A}_{\mathcal{I}_i,:}}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}} \right)^{\mathrm{T}} \frac{\mathbf{A}_{\mathcal{I}_i,:}}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}} (\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}) \right\|_2^2 \\
&\le \|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2 - \left(2\alpha - \frac{\alpha^2 \sigma_{\max}^2(\mathbf{A}_{\mathcal{I}_i,:})}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2} \right) \frac{\|\mathbf{A}_{\mathcal{I}_i,:}(\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b})\|_2^2}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2} \\
&\qquad \text{(by Lemma 2.5)} \\
&\le \|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2 - \frac{(2\alpha - \alpha^2 \beta_{\max}^{\mathcal{I}}) \|\mathbf{A}_{\mathcal{I}_i,:}(\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b})\|_2^2}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2},
\end{aligned}$$

we have

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] &\le \|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2 - \frac{(2\alpha - \alpha^2 \beta_{\max}^{\mathcal{I}}) \|\mathbf{A}(\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b})\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} \\
&\le \|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2 - \frac{(2\alpha - \alpha^2 \beta_{\max}^{\mathcal{I}}) \sigma_{\min}^2(\mathbf{A}) \|(\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b})\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} \\
&\qquad \text{(by Lemma 2.4 and } 0 < \alpha < 2/\beta_{\max}^{\mathcal{I}}) \\
&= \eta \|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2,
\end{aligned}$$

which yields

$$\mathbb{E}\left[\|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] \le \eta \mathbb{E}\left[\|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right]. \tag{2.4}$$

Note that for any $\varepsilon > 0$, we have

$$\begin{aligned}
\|\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2 &= \|\mathbf{x}^k - \widehat{\mathbf{x}}^k + \widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2 \\
&\le (\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2 + \|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2)^2 \\
&\le \|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2 + \|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2 + 2\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2 \|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2 \\
&\le \left(1 + \frac{1}{\varepsilon}\right) \|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2 + (1+\varepsilon) \|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2. \qquad (2.5)
\end{aligned}$$

Combining (2.3), (2.4), and (2.5) yields

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] &\le \left(1 + \frac{1}{\varepsilon}\right) \mathbb{E}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right] + (1+\varepsilon) \mathbb{E}\left[\|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] \\
&\le \left(1 + \frac{1}{\varepsilon}\right) \frac{\alpha^2 \beta_{\max}^{\mathcal{I}} \rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2} \|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2 + (1+\varepsilon)\eta \mathbb{E}\left[\|\mathbf{x}^{k-1} - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] \\
&\le \left(1 + \frac{1}{\varepsilon}\right) \frac{\alpha^2 \beta_{\max}^{\mathcal{I}} \|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} (\rho^k + \rho^{k-1}(1+\varepsilon)\eta) \\
&\quad + (1+\varepsilon)^2 \eta^2 \mathbb{E}\left[\|\mathbf{x}^{k-2} - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] \\
&\le \cdots \\
&\le \left(1 + \frac{1}{\varepsilon}\right) \frac{\alpha^2 \beta_{\max}^{\mathcal{I}} \|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} \sum_{l=0}^{k-1} \rho^{k-l} (1+\varepsilon)^l \eta^l \\
&\quad + (1+\varepsilon)^k \eta^k \|\mathbf{x}^0 - \mathbf{A}^\dagger \mathbf{b}\|_2^2.
\end{aligned}$$

This completes the proof. ∎

Remark 2.8. For the case REABK with $s = m$, $t = n$, and $\alpha = 1$ (i.e., REK), we have

$$\beta_{\max}^{\mathcal{I}} = \max_{i\in[m]} \frac{\|\mathbf{A}_{i,:}\|_2^2}{\|\mathbf{A}_{i,:}\|_{\mathrm{F}}^2} = 1, \quad \beta_{\max}^{\mathcal{J}} = \max_{j\in[n]} \frac{\|\mathbf{A}_{:,j}\|_2^2}{\|\mathbf{A}_{:,j}\|_{\mathrm{F}}^2} = 1.$$

Therefore,

$$\eta = 1 - \frac{(2\alpha - \alpha^2\beta_{\max}^{\mathcal{I}})\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2} = 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}$$

and

$$\rho = 1 - \frac{(2\alpha - \alpha^2\beta_{\max}^{\mathcal{J}})\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2} = 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}.$$

It follows from

$$\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b} = \left(\mathbf{I} - \frac{(\mathbf{A}_{i,:})^{\mathrm{T}}\mathbf{A}_{i,:}}{\|\mathbf{A}_{i,:}\|_2^2}\right)(\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b})$$

and

$$\mathbf{x}^k - \widehat{\mathbf{x}}^k = \frac{\mathbf{b}_i - \mathbf{A}_{i,:}\mathbf{A}^\dagger\mathbf{b} - \mathbf{z}_i^k}{\|\mathbf{A}_{i,:}\|_2^2}(\mathbf{A}_{i,:})^{\mathrm{T}}$$

that

$$(\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b})^{\mathrm{T}}(\mathbf{x}^k - \widehat{\mathbf{x}}^k) = 0.$$

Then we have

$$\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2 = \|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2 + \|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2,$$

which yields the following convergence for REK (see [13]):

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] &= \mathbb{E}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right] + \mathbb{E}\left[\|\widehat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\le \frac{\alpha^2\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2 + \rho\mathbb{E}\left[\|\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\le \frac{2\alpha^2\rho^k\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} + \rho^2\mathbb{E}\left[\|\mathbf{x}^{k-2} - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\le \cdots \\
&\le \rho^k\left(\frac{k\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2} + \|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right).
\end{aligned}$$

Actually our proof is a modification of that of Zouzias and Freris [52]. We reorganize the arguments used by Zouzias and Freris and refine the analysis to get a better convergence estimate.

Remark 2.9. Let $\widehat{\rho} := \max(\eta, \rho)$ and $\beta_{\max} := \max(\beta_{\max}^{\mathcal{I}}, \beta_{\max}^{\mathcal{J}})$. Then we have

$$\widehat{\rho} = 1 - \frac{(2\alpha - \alpha^2\beta_{\max})\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}.$$

By Theorem 2.7, we have

$$
\begin{aligned}
&\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\leq (1+\varepsilon)^k\eta^k\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2 + \left(1+\frac{1}{\varepsilon}\right)\frac{\alpha^2\beta_{\max}^{\mathcal{I}}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\sum_{l=0}^{k-1}\rho^{k-l}(1+\varepsilon)^l\eta^l \\
&\leq (1+\varepsilon)^k\widehat{\rho}^k\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2 + \left(1+\frac{1}{\varepsilon}\right)\frac{\alpha^2\beta_{\max}^{\mathcal{I}}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\widehat{\rho}^k\sum_{l=0}^{k-1}(1+\varepsilon)^l \\
&\leq (1+\varepsilon)^k\widehat{\rho}^k\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2 + \left(1+\frac{1}{\varepsilon}\right)\frac{\alpha^2\beta_{\max}^{\mathcal{I}}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\widehat{\rho}^k\frac{(1+\varepsilon)^k - 1}{\varepsilon} \\
&\leq (1+\varepsilon)^k\widehat{\rho}^k\left(\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2 + \frac{(1+\varepsilon)\alpha^2\beta_{\max}^{\mathcal{I}}\|\mathbf{z}^0 - \mathbf{b}_\perp\|_2^2}{\varepsilon^2\|\mathbf{A}\|_{\mathrm{F}}^2}\right),
\end{aligned}
$$

which shows that REABK exponentially converges in the mean square to the minimum norm least squares solution of a given linear system of equations with the rate $(1+\varepsilon)\widehat{\rho}$ if $0 < \alpha < 2/\beta_{\max}$. Setting $\alpha = 1/\beta_{\max}$ yields

$$\widehat{\rho} = 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\beta_{\max}\|\mathbf{A}\|_{\mathrm{F}}^2},$$

which is better than the rate of REK (see Remark 2.8)

$$\rho = 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}$$

if $\beta_{\max} < 1$. The smaller $\beta_{\max}$, the faster the convergence of REABK in terms of the numbers of iterations. We give the possible lower bound for $\beta_{\max}$ in the following. Recalling that

$$\beta_{\max}^{\mathcal{I}} = \max_{i\in[s]}\frac{\sigma_{\max}^2(\mathbf{A}_{\mathcal{I}_i,:})}{\|\mathbf{A}_{\mathcal{I}_i,:}\|_{\mathrm{F}}^2} \quad \text{and} \quad \beta_{\max}^{\mathcal{J}} = \max_{j\in[t]}\frac{\sigma_{\max}^2(\mathbf{A}_{:,\mathcal{J}_j})}{\|\mathbf{A}_{:,\mathcal{J}_j}\|_{\mathrm{F}}^2},$$

we have

$$\max_{i\in[s]}\frac{1}{|\mathcal{I}_i|} \leq \max_{i\in[s]}\frac{1}{\operatorname{rank}(\mathbf{A}_{\mathcal{I}_i,:})} \leq \beta_{\max}^{\mathcal{I}} \leq 1$$

and

$$\max_{j\in[t]}\frac{1}{|\mathcal{J}_j|} \leq \max_{j\in[t]}\frac{1}{\operatorname{rank}(\mathbf{A}_{:,\mathcal{I}_j})} \leq \beta_{\max}^{\mathcal{J}} \leq 1.$$

Therefore,

$$\max\left(\max_{i\in[s]}\frac{1}{|\mathcal{I}_i|}, \max_{j\in[t]}\frac{1}{|\mathcal{J}_j|}\right) \leq \beta_{\max} \leq 1,$$

which means that REABK is at least as fast as REK in terms of iterations. The numerical results in section 3 show that the convergence of REABK with appropriate block sizes and stepsizes is much faster than that of REK both in the numbers of iterations and the computing times.

Remark 2.10. It was shown in [20] that the convergence of $\mathbf{x}^k$ to $\mathbf{A}^\dagger\mathbf{b}$ under the expected norm of the error (Theorem 2.7) is a stronger form of convergence than the convergence of the norm of the expected error (Theorem 2.2), as the former also guarantees that the variance of $\mathbf{x}_i^k$ (the ith element of $\mathbf{x}^k$) converges to zero for

$i = 1, \ldots, n$. By Remark 2.3, we know that $0 < \alpha < 2\|\mathbf{A}\|_{\mathrm{F}}^2/\sigma_{\max}^2(\mathbf{A})$ guarantees the convergence of the norm of the expected error. By Remark 2.9, we know that $0 < \alpha < 2/\beta_{\max}$ guarantees the convergence of the expected norm of the error. However, since the convergence estimate in Remark 2.9 usually is not sharp, the stepsize α satisfying $2/\beta_{\max} \leq \alpha < 2\|\mathbf{A}\|_{\mathrm{F}}^2/\sigma_{\max}^2(\mathbf{A})$ is also possible to result in convergence (see Figure 1 and Tables 2 and 3 in section 3).

3. Numerical results. In this section, we compare the performance of the REABK algorithm proposed in this paper against the REK algorithm [52] and the projection-based RDBK algorithm [41] on a variety of test problems. We do not claim optimized implementations of the algorithms and only run on small- or medium-scale problems. The purpose is only to demonstrate that even in these simple examples, REABK offers significant advantages over REK. All experiments are performed using MATLAB (version R2019a) on a laptop with a 2.7-GHz Intel Core i7 processor, 16 GB memory, and a Mac operating system.

To construct an inconsistent linear system, we set $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{r}$, where $\mathbf{x}$ is a vector with entries generated from a standard normal distribution and the residual $\mathbf{r} \in \mathrm{null}(\mathbf{A}^{\mathrm{T}})$. Note that one can obtain such a vector $\mathbf{r}$ by the MATLAB function `null`. For all algorithms, we set $\mathbf{z}^0 = \mathbf{b}$ and $\mathbf{x}^0 = \mathbf{0}$ and stop if the error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2 \leq 10^{-5}$. We report the average number of iterations (denoted as ITER) and the average computing time in seconds (denoted as CPU) of REK, RDBK, and REABK. Note that `A\b` will usually not be the same as `pinv(A)*b` when `A` is rank-deficient or underdetermined. We use MATLAB's `lsqminnorm` (which is typically more efficient than `pinv`) to solve the small least squares problems at each step of RDBK. We refer the reader to [40, 41] for more numerical aspects of RDBK. We also report the speedup of REABK against REK, which is defined as

$$\text{speedup} = \frac{\text{CPU of REK}}{\text{CPU of REABK}}.$$

For the block methods, we assume that the subsets $\{\mathcal{I}_i\}_{i=1}^{s-1}$ and $\{\mathcal{J}_j\}_{j=1}^{t-1}$ have the same size τ (i.e., $|\mathcal{I}_i| = |\mathcal{J}_j| = \tau$). We consider the row partition $\{\mathcal{I}_i\}_{i=1}^{s}$,

$$\begin{aligned}
\mathcal{I}_i &= \{(i-1)\tau + 1, (i-1)\tau + 2, \ldots, i\tau\}, \quad i = 1, 2, \ldots, s-1, \\
\mathcal{I}_s &= \{(s-1)\tau + 1, (s-1)\tau + 2, \ldots, m\}, \quad |\mathcal{I}_s| \leq \tau,
\end{aligned}$$

and the column partition $\{\mathcal{J}_j\}_{j=1}^{t}$,

$$\begin{aligned}
\mathcal{J}_j &= \{(j-1)\tau + 1, (j-1)\tau + 2, \ldots, j\tau\}, \quad j = 1, 2, \ldots, t-1, \\
\mathcal{J}_t &= \{(t-1)\tau + 1, (t-1)\tau + 2, \ldots, n\}, \quad |\mathcal{J}_t| \leq \tau.
\end{aligned}$$

3.1. Synthetic data. Two types of coefficient matrices are generated as follows.

- Type I: For given m, n, $r = \mathrm{rank}(\mathbf{A})$, and $\kappa > 1$, we construct a matrix $\mathbf{A}$ by

  $$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}},$$

  where $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$. Entries of $\mathbf{U}$ and $\mathbf{V}$ are generated from a standard normal distribution, and then columns are orthonormalized:

  $$[\mathtt{U}, \sim] = \mathtt{qr}(\mathtt{randn}(\mathtt{m}, \mathtt{r}), 0); \qquad [\mathtt{V}, \sim] = \mathtt{qr}(\mathtt{randn}(\mathtt{n}, \mathtt{r}), 0).$$

  The matrix $\mathbf{D}$ is an $r \times r$ diagonal matrix whose diagonal entries are uniformly

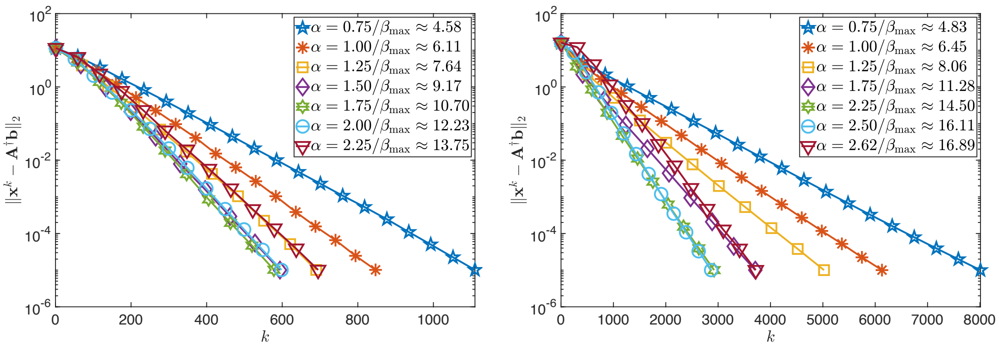

FIG. 1. *The average (10 trials of each case) error* $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2$ *of* REABK *with block size* $\tau = 10$ *and different stepsizes* α *from* $0.75/\beta_{\max}$ *to* $2.62/\beta_{\max}$ *for two inconsistent linear systems. Left: Type I matrix* $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$ *with* $m = 500$*,* $n = 250$*,* $r = 150$*,* $\kappa = 2$*. Right: Type II matrix* $\mathbf{A} = \texttt{randn}(500, 250)$*.*

distributed numbers in $(1, \kappa)$:

$$\mathtt{D} = \mathtt{diag}(\mathtt{1} + (\kappa - \mathtt{1}).*\mathtt{rand}(\mathtt{r}, \mathtt{1})).$$

So the condition number of $\mathbf{A}$ is upper bounded by κ.

- Type II: For given m, n, entries of $\mathbf{A}$ are generated from a standard normal distribution:

$$\mathtt{A} = \mathtt{randn}(\mathtt{m}, \mathtt{n}).$$

So $\mathbf{A}$ is a full-rank matrix almost surely.

In Figure 1, we plot the error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2$ of REABK with a fixed block size ($\tau = 10$) and different stepsizes (α from $0.75/\beta_{\max}$ to $2.62/\beta_{\max}$) for two inconsistent linear systems with coefficient matrices of Types I ($\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$ with $m = 500$, $n = 250$, $r = 150$, $\kappa = 2$) and II ($\mathbf{A} = \texttt{randn(500,250)}$). It is observed that the convergence of REABK becomes faster with an increase in stepsize and then slows down after reaching the fastest rate.

In Tables 1 and 2, we report the numbers of iterations and the computing times of the REK, RDBK, and REABK algorithms for solving inconsistent linear systems. For the block algorithms (RDBK and REABK), a fixed block size $\tau = 10$ is used. For the REABK algorithm, empirical stepsizes $\alpha = 1.75/\beta_{\max}$ and $\alpha = 2.25/\beta_{\max}$ are used for Type I and Type II matrices, respectively. From these two tables, we observe (i) that the RDBK and REABK algorithms vastly outperform the REK algorithm in terms of both the numbers of iterations and the computing times and (ii) that the convergence rates of the RDBK and REABK algorithms are almost the same (compared with REK) in terms of the numbers of iterations and that the REABK algorithm is faster than the RDBK algorithm in terms of the computing time due to the pseudoinverse-free nature (which means less computing cost per iteration).

In Figure 2, we plot the error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2$ and the computing times of REABK with block sizes $\tau = 5, 10, 20, 50, 100, 200$ and stepsize $\alpha = 1.75/\beta_{\max}$ for two inconsistent linear systems with coefficient matrices of Types I ($\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$ with $m = 20000$, $n = 5000$, $r = 4500$, $\kappa = 2$) and II ($\mathbf{A} = \texttt{randn(20000,5000)}$). The average numbers of required iterations are also reported. We observe (i) that increasing block size and using the empirical stepsize $\alpha = 1.75/\beta_{\max}$ lead to a better convergence in terms of the numbers of iterations and (ii) that with the increase of block size, the computing time first decreases, then increases after reaching the minimum value, and finally

Table 1
The average (10 trials of each algorithm) ITER *and* CPU *of* REK, RDBK*($\tau = 10$), and* REABK*($\tau = 10$, $\alpha = 1.75/\beta_{\max}$) for inconsistent linear systems with random coefficient matrices* $\mathbf{A}$ *of Type* I: $\mathbf{A} = \mathbf{UDV}^{\mathrm{T}}$.

$m \times n$	rank	κ	REK		RDBK		REABK			
			ITER	CPU	ITER	CPU	α	ITER	CPU	speedup
250×500	150	2	5826	0.26	572	0.21	10.87	586	0.05	4.90
250×500	150	10	65520	2.87	6166	2.19	9.36	7365	0.63	4.59
500×1000	250	2	10068	0.59	1000	0.43	11.82	991	0.13	4.60
500×1000	250	10	114297	6.61	10209	4.29	10.85	10259	1.23	5.36
500×250	150	2	5755	0.25	562	0.19	10.70	578	0.03	7.32
500×250	150	10	63741	2.76	5784	1.90	10.13	6424	0.36	7.60
500×250	250	2	9971	0.43	940	0.31	12.47	961	0.06	7.81
500×250	250	10	119182	5.14	11328	3.73	10.99	10783	0.61	8.43
1000×500	250	2	9959	0.55	974	0.39	12.10	987	0.10	5.53
1000×500	250	10	118134	6.54	11236	4.44	11.20	10349	1.03	6.36
1000×500	500	2	20188	1.11	2007	0.80	13.84	2115	0.21	5.20
1000×500	500	10	254117	14.01	25361	10.00	12.67	20432	2.03	6.92

Table 2
The average (10 trials of each algorithm) ITER *and* CPU *of* REK, RDBK*($\tau = 10$), and* REABK*($\tau = 10$, $\alpha = 2.25/\beta_{\max}$) for inconsistent linear systems with random coefficient matrices* $\mathbf{A}$ *of Type* II: $\mathbf{A} =$ `randn(m,n)`.

$m \times n$	rank	$\frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}$	REK		RDBK		REABK			
			ITER	CPU	ITER	CPU	α	ITER	CPU	speedup
250×120	120	5.25	18060	0.66	1646	0.50	13.48	1337	0.06	10.54
500×250	250	5.73	41016	1.79	3811	1.26	14.50	2885	0.17	10.81
750×370	370	5.80	59660	2.91	5929	2.20	16.23	4115	0.36	8.07
1000×500	500	5.74	83093	4.61	8183	3.25	16.42	5422	0.55	8.41

tends to be stable. This means that for sufficiently large block size, the decrease in iteration complexity cannot compensate for the increase in cost per iteration. On the other hand, if the REABK algorithm is implemented in parallel, a larger τ will be better.

In Figure 3, we plot the computing times of the REK, RDBK, and REABK algorithms for inconsistent linear systems with coefficient matrices of Types I ($\mathbf{A} = \mathbf{UDV}^{\mathrm{T}}$ with $m = 2000, 4000, \ldots, 20000$, $n = 500$, $r = 250$, $\kappa = 2$) and II ($\mathbf{A} =$ `randn(m,n)` with $m = 2000, 4000, \ldots, 20000$, $n = 500$). For all cases, the block size $\tau = 10$ and the stepsize $\alpha = 1.75/\beta_{\max}$ are used. We observe that both RDBK and REABK are better than REK and that REABK is the best.

3.2. Real-world data. Finally, we test REK, RDBK, and REABK using eight inconsistent linear systems with coefficient matrices from the University of Florida sparse matrix collection [10]. The eight matrices are `abtaha1`, `flower_5_1`, `football`, `lp_nug15`, `relat6`, `relat7`, `Sandi_authors`, and `WorldCities`. In Table 3, we report the numbers of iterations and the computing times for the REK, RDBK, and REABK algorithms. For each matrix, we tested two stepsizes of REABK: The first is $1/\beta_{\max}$, and the second is some value in the open interval $(1/\beta_{\max}, 2\|\mathbf{A}\|_{\mathrm{F}}^2/\sigma_{\max}^2(\mathbf{A}))$. We observe that REABK based on good choices of block size and stepsize significantly outperforms REK. Moreover, good stepsize and block size are problem dependent. In all cases, RDBK requires the least number of iterations. On the case of `relat7`, RDBK is the winner. On the cases of `flower_5_1`, `lp_nug15`, and `WorldCities`,

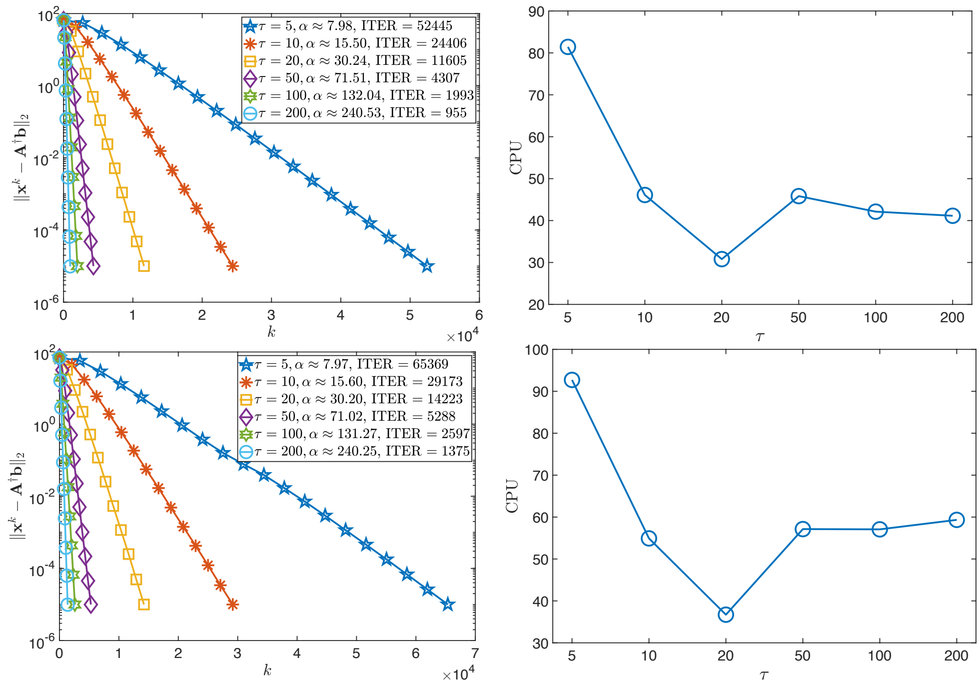

FIG. 2. *The average (10 trials of each case) error* $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2$ *and CPU of REABK with different block sizes* $\tau = 5, 10, 20, 50, 100, 200$ *and stepsize* $\alpha = 1.75/\beta_{\max}$ *for inconsistent linear systems. The average numbers of required iterations are also reported. Upper: Type* I *matrix* $\mathbf{A} = \mathbf{UDV}^{\mathrm{T}}$ *with* $m = 20000$, $n = 5000$, $r = 4500$, *and* $\kappa = 2$. *Lower: Type* II *matrix* $\mathbf{A} =$ `randn(20000, 5000)`.

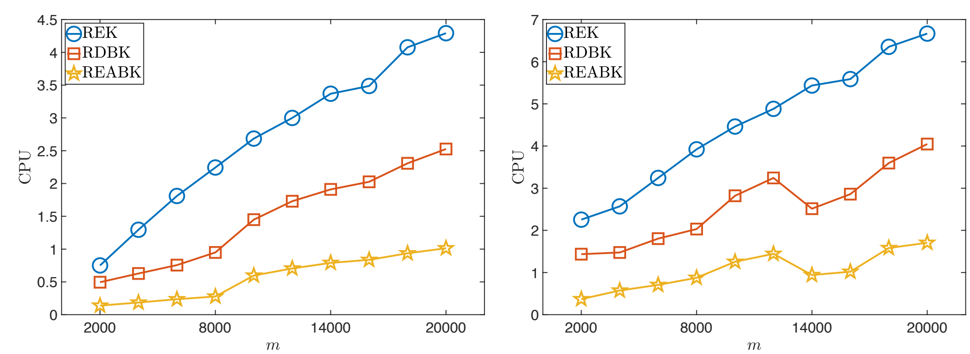

FIG. 3. *The average (10 trials of each algorithm) CPU of REK, RDBK(*$\tau = 10$*), and REABK(*$\tau = 10, \alpha = 1.75/\beta_{\max}$*) for inconsistent linear systems. Left: Type* I *matrix* $\mathbf{A} = \mathbf{UDV}^{\mathrm{T}}$ *with* $m = 2000, 4000, \ldots, 20000$, $n = 500$, $r = 250$, *and* $\kappa = 2$. *Right: Type* II *matrix* $\mathbf{A} =$ `randn(m,n)` *with* $m = 2000, 4000, \ldots, 20000$ *and* $n = 500$.

RDBK requires the most computing time due to the decrease in iteration complexity, which cannot compensate for the increase in cost per iteration.

4. Concluding remarks. We have proposed an REABK algorithm for solving general linear systems and prove its convergence theory. At each step, REABK uses two RABK (with special choice of weights) updates. The new algorithm can be implemented for parallel computation. Numerical experiments show that the crucial

Table 3

The average (10 trials of each algorithm) ITER *and* CPU *of* REK, RDBK(τ), *and* REABK(τ, α) *for inconsistent linear systems with coefficient matrices from* [10]. *For each matrix, two stepsizes of* REABK *are tested: The first is* $1/\beta_{\max}$, *and the second is some value in the open interval* $(1/\beta_{\max}, 2\|\mathbf{A}\|_F^2/\sigma_{\max}^2(\mathbf{A}))$.

Matrix	$m \times n$	rank	$\frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}$	REK		τ	RDBK		REABK			
				ITER	CPU		ITER	CPU	α	ITER	CPU	speedup
abtaha1	14596×209	209	12.23	276946	89.38	10	30385	43.98	1.82	151395	68.30	1.31
									5	56064	25.34	3.53
flower_5_1	211×201	179	13.70	135117	5.16	5	24297	6.22	1	136037	6.15	0.84
									4	34381	1.55	3.34
football	35×35	19	166.47	810792	21.99	5	89141	23.78	1	858215	30.64	0.72
									2	409995	14.63	1.50
lp_nug15	6330×22275	5698	2.73	216924	220.64	20	26363	253.79	3.53	40539	199.67	1.10
									5	31039	158.29	1.39
relat6	2340×157	137	7.74	34536	2.43	10	3797	1.76	1	34273	3.81	0.64
									2.5	13971	1.56	1.56
relat7	21924×1045	1012	10.85	550810	283.69	10	67445	153.42	1	542100	466.89	0.61
									2.5	218287	188.81	1.50
Sandi_authors	86×86	72	189.58	2525141	73.28	5	141826	34.66	1	2533343	99.36	0.74
									2.5	999294	39.15	1.87
WorldCities	315×100	100	66.00	120699	4.32	5	18160	4.69	1.13	105647	4.52	0.96
									2.5	47372	2.02	2.14

point for guaranteeing fast convergence is to obtain good block size and stepsize. Finding appropriate variable stepsize by the adaptive extrapolation [36] and proposing more effective partitions (including the corresponding probability distributions) based on the techniques of [40, 12, 50, 36] should be valuable topics. We also note that RABK allows the flexibility that the distributions from which blocks are selected do not require the blocks to form a partition of the columns or rows. Designing variants of REABK based on RABK with random samplings that do not depend on the partitions is straightforward. We believe the technique used in the proof of Theorem 2.7 still works for these variants, although the analysis will be more complicated. In addition, developing parallel and accelerated variants of REABK based on the approach used by Richtárik and Takáč [48] is also worth exploring. We will work on these topics in the future.

Acknowledgments. The authors are thankful to the referees for their detailed comments and valuable suggestions that have led to remarkable improvements.

REFERENCES

[1] M. Arioli, I. Duff, J. Noailles, and D. Ruiz, *A block projection method for sparse matrices*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 47–70, https://doi.org/10.1137/0913003, https://mathscinet.ams.org/mathscinet-getitem?mr=1145175.

[2] Z.-Z. Bai and W.-T. Wu, *On convergence rate of the randomized Kaczmarz method*, Linear Algebra Appl., 553 (2018), pp. 252–269, https://doi.org/10.1016/j.laa.2018.05.009.

[3] Z.-Z. Bai and W.-T. Wu, *On greedy randomized Kaczmarz method for solving large sparse linear systems*, SIAM J. Sci. Comput., 40 (2018), pp. A592–A606, https://doi.org/10.1137/17M1137747.

[4] Z.-Z. Bai and W.-T. Wu, *On relaxed greedy randomized Kaczmarz methods for solving large sparse linear systems*, Appl. Math. Lett., 83 (2018), pp. 21–26, https://doi.org/10.1016/j.aml.2018.03.008.

[5] Z.-Z. Bai and W.-T. Wu, *On greedy randomized coordinate descent methods for solving large linear least-squares problems*, Numer. Linear Algebra Appl., 26 (2019), pp. e2237, 15, https://doi.org/10.1002/nla.2237.

[6] Z.-Z. Bai and W.-T. Wu, *On partially randomized extended Kaczmarz method for solving large sparse overdetermined inconsistent linear systems*, Linear Algebra Appl., 578 (2019), pp. 225–250, https://doi.org/10.1016/j.laa.2019.05.005.

[7] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC 15, Springer-Verlag, New York, 2003, https://mathscinet.ams.org/mathscinet-getitem?mr=1987382.

[8] R. Bramley and A. Sameh, *Row projection methods for large nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 168–193, https://doi.org/10.1137/0913010.

[9] C. Byrne, *A unified treatment of some iterative algorithms in signal processing and image reconstruction*, Inverse Problems, 20 (2004), pp. 103–120, https://doi.org/10.1088/0266-5611/20/1/006.

[10] T. A. Davis and Y. Hu, *The University of Florida sparse matrix collection*, ACM Trans. Math. Software, 38 (2011), pp. Art. 1, 25, https://doi.org/10.1145/2049662.2049663.

[11] J. A. De Loera, J. Haddock, and D. Needell, *A sampling Kaczmarz-Motzkin algorithm for linear feasibility*, SIAM J. Sci. Comput., 39 (2017), pp. S66–S87, https://doi.org/10.1137/16M1073807.

[12] L. A. Drummond, I. S. Duff, R. Guivarch, D. Ruiz, and M. Zenadi, *Partitioning strategies for the block Cimmino algorithm*, J. Engrg. Math., 93 (2015), pp. 21–39, https://doi.org/10.1007/s10665-014-9699-0.

[13] K. Du, *Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss-Seidel algorithms*, Numer. Linear Algebra Appl., 26 (2019), pp. e2233, 14, https://doi.org/10.1002/nla.2233.

[14] K. Du and H. Gao, *A new theoretical estimate for the convergence rate of the maximal weighted residual Kaczmarz algorithm*, Numer. Math. Theory Methods Appl., 12 (2019), pp. 627–639, https://mathscinet.ams.org/mathscinet-getitem?mr=3894847.

[15] K. Du and X. Sun, *A Doubly Stochastic Block Gauss-Seidel Algorithm for Solving Linear Equations*, preprint, arXiv:1912.13291, 2019.

[16] I. S. Duff, R. Guivarch, D. Ruiz, and M. Zenadi, *The augmented block Cimmino distributed method*, SIAM J. Sci. Comput., 37 (2015), pp. A1248–A1269, https://doi.org/10.1137/140961444.

[17] B. Dumitrescu, *On the relation between the randomized extended Kaczmarz algorithm and coordinate descent*, BIT, 55 (2015), pp. 1005–1015, https://doi.org/10.1007/s10543-014-0526-9.

[18] T. Elfving, *Block-iterative methods for consistent and inconsistent linear equations*, Numer. Math., 35 (1980), pp. 1–12, https://doi.org/10.1007/BF01396365.

[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013.

[20] R. M. Gower and P. Richtárik, *Randomized iterative methods for linear systems*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1660–1690, https://doi.org/10.1137/15M1025487.

[21] J. Haddock and D. Needell, *On Motzkin's method for inconsistent linear systems*, BIT, 59 (2019), pp. 387–401, https://doi.org/10.1007/s10543-018-0737-6.

[22] J. Haddock and D. Needell, *Randomized projection methods for linear systems with arbitrarily large sparse corruptions*, SIAM J. Sci. Comput., 41 (2019), pp. S19–S36, https://doi.org/10.1137/18M1179213.

[23] A. Hefny, D. Needell, and A. Ramdas, *Rows versus columns: Randomized Kaczmarz or Gauss-Seidel for ridge regression*, SIAM J. Sci. Comput., 39 (2017), pp. S528–S542, https://doi.org/10.1137/16M1077891.

[24] G. T. Herman, *Fundamentals of Computerized Tomography*, 2nd ed., Advances in Pattern Recognition, Image Reconstruction from Projections, Springer, Dordrecht, the Netherlands, 2009, https://doi.org/10.1007/978-1-84628-723-7.

[25] G. T. Herman and R. Davidi, *Image reconstruction from a small number of projections*, Inverse Problems, 24 (2008), 045011, https://doi.org/10.1088/0266-5611/24/4/045011.

[26] G. T. Herman and L. B. Meyer, *Algebraic reconstruction techniques can be made computationally efficient*, IEEE Trans. Med. Imaging, 12 (1993), pp. 600–609.

[27] S. Kaczmarz, *Angenäherte auflösung von systemen linearer gleichungen*, Bull. Intern. Acad. Polonaise Sci. Lett., Cl. Sci. Math. Nat. A, 35 (1937), pp. 355–357.

[28] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Classics in Applied Mathematics 33, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898719277, reprint of the 1988 original.

[29] D. Leventhal and A. S. Lewis, *Randomized methods for linear constraints: Convergence rates and conditioning*, Math. Oper. Res., 35 (2010), pp. 641–654, https://doi.org/10.1287/moor.1100.0456.

[30] J. Liu and S. J. Wright, *An accelerated randomized Kaczmarz algorithm*, Math. Comp., 85 (2016), pp. 153–178, https://doi.org/10.1090/mcom/2971.

[31] Y. Liu and C.-Q. Gu, *Variant of greedy randomized Kaczmarz for ridge regression*, Appl. Numer. Math., 143 (2019), pp. 223–246, https://doi.org/10.1016/j.apnum.2019.04.008.

[32] D. A. Lorenz, S. Wenger, F. Schöpfer, and M. Magnor, *A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing*, in IEEE International Conference on Image Processing, 2015, pp. 1347 – 1351.

[33] A. Ma, D. Needell, and A. Ramdas, *Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1590–1604, https://doi.org/10.1137/15M1014425.

[34] J. D. Moorman, T. K. Tu, D. Molitor, and D. Needell, *Randomized Kaczmarz with Averaging*, preprint, arXiv:2002.04126, 2020.

[35] F. Natterer, *The Mathematics of Computerized Tomography*, Classics in Applied Mathematics 32, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898719284.

[36] I. Necoara, *Faster randomized block Kaczmarz algorithms*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 1425–1452, https://doi.org/10.1137/19M1251643.

[37] I. Necoara, P. Richtárik, and A. Patrascu, *Randomized projection methods for convex feasibility: Conditioning and convergence rates*, SIAM J. Optim., 29 (2019), pp. 2814–2852, https://doi.org/10.1137/18M1167061.

[38] D. Needell, *Randomized Kaczmarz solver for noisy linear systems*, BIT, 50 (2010), pp. 395–403, https://doi.org/10.1007/s10543-010-0265-5.

[39] D. Needell, N. Srebro, and R. Ward, *Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm*, Math. Program., 155 (2016), pp. 549–573, https://doi.org/10.1007/s10107-015-0864-7.

[40] D. Needell and J. A. Tropp, *Paved with good intentions: Analysis of a randomized block Kaczmarz method*, Linear Algebra Appl., 441 (2014), pp. 199–221, https://doi.org/10.1016/j.laa.2012.12.022.

[41] D. Needell, R. Zhao, and A. Zouzias, *Randomized block Kaczmarz method with projection for solving least squares*, Linear Algebra Appl., 484 (2015), pp. 322–343, https://doi.org/10.1016/j.laa.2015.06.027.

[42] S. Petra and C. Popa, *Single projection Kaczmarz extended algorithms*, Numer. Algorithms, 73 (2016), pp. 791–806, https://doi.org/10.1007/s11075-016-0118-7.

[43] B. T. Poljak, *Gradient methods for minimizing functionals*, Ž. Vyčisl. Mat. Mat. Fiz., 3 (1963), pp. 643–653.

[44] C. Popa, *Extensions of block-projections methods with relaxation parameters to inconsistent and rank-deficient least-squares problems*, BIT, 38 (1998), pp. 151–176, https://doi.org/10.1007/BF02510922.

[45] C. Popa, *Convergence rates for Kaczmarz-type algorithms*, Numer. Algorithms, 79 (2018), pp. 1–17, https://doi.org/10.1007/s11075-017-0425-7.

[46] C. Popa and R. Zdunek, *Kaczmarz extended algorithm for tomographic image reconstruction from limited data*, Math. Comput. Simulation, 65 (2004), pp. 579–598, https://doi.org/10.1016/j.matcom.2004.01.021.

[47] M. Razaviyayn, M. Hong, N. Reyhanian, and Z.-Q. Luo, *A linearly convergent doubly stochastic Gauss-Seidel algorithm for solving linear equations and a certain class of over-parameterized optimization problems*, Math. Program., 176 (2019), pp. 465–496, https://doi.org/10.1007/s10107-019-01404-0.

[48] P. Richtárik and M. Takáč, *Stochastic reformulations of linear systems algorithms and convergence theory*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 487–524.

[49] T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence*, J. Fourier Anal. Appl., 15 (2009), pp. 262–278, https://doi.org/10.1007/s00041-008-9030-4.

[50] F. S. Torun, M. Manguoglu, and C. Aykanat, *A novel partitioning method for accelerating the block Cimmino algorithm*, SIAM J. Sci. Comput., 40 (2018), pp. C827–C850, https://doi.org/10.1137/18M1166407.

[51] J.-J. Zhang, *A new greedy Kaczmarz algorithm for the solution of very large linear systems*, Appl. Math. Lett., 91 (2019), pp. 207–212, https://doi.org/10.1016/j.aml.2018.12.022.

[52] A. Zouzias and N. M. Freris, *Randomized extended Kaczmarz for solving least squares*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 773–793, https://doi.org/10.1137/120889897.