Received: 17 January 2018 | Revised: 16 September 2018 | Accepted: 18 January 2019
DOI: 10.1002/nla.2233

Check for updates

RESEARCH ARTICLE

WILEY

Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss–Seidel algorithms

Kui Du

Fujian Provincial Key Laboratory of Mathematical Modeling and High-Performance Scientific Computing, School of Mathematical Sciences, Xiamen University, Xiamen, China

Correspondence
Kui Du, Fujian Provincial Key Laboratory of Mathematical Modeling and High-Performance Scientific Computing, School of Mathematical Sciences, Xiamen University, Xiamen 361005, China.
Email: kuidu@xmu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 11771364 and 91430213; Fundamental Research Funds for the Central Universities, Grant/Award Number: 20720160002

Summary

The randomized extended Kaczmarz and Gauss–Seidel algorithms have attracted much attention because of their ability to treat all types of linear systems (consistent or inconsistent, full rank or rank deficient). In this paper, we present *tight* upper bounds for the convergence of the randomized extended Kaczmarz and Gauss–Seidel algorithms. Numerical experiments are given to illustrate the theoretical results.

KEYWORDS

convergence analysis, Moore–Penrose pseudoinverse solution, randomized extended Gauss–Seidel algorithm, randomized extended Kaczmarz algorithm, tight upper bound

1 | INTRODUCTION

Due to the better performance in many situations than existing classical iterative algorithms, randomized iterative algorithms for solving a linear system of equations

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

have attracted much attention recently; see, for example, other works[1–19] and the references therein. In this paper, we consider the randomized Kaczmarz (RK) algorithm,[1] the randomized Gauss–Seidel (RGS) algorithm,[2] the randomized extended Kaczmarz (REK) algorithm,[5] and the randomized extended Gauss–Seidel (REGS) algorithm.[11] Convergence rates of these algorithms (including their deterministic variants) have been considered extensively; see, for example, other works.[1,2,5,6,11,13,20–26] Let $\mathbf{A}^\dagger$ denote the Moore–Penrose pseudoinverse[27] of $\mathbf{A}$. We summarize the convergence of RK, RGS, REK, and REGS in expectation to the Moore–Penrose pseudoinverse solution $\mathbf{A}^\dagger\mathbf{b}$ for all types of linear systems in Table 1.

Main contributions. We present *tight* upper bounds (in the sense that there is a linear system $\mathbf{Ax} = \mathbf{b}$ for which the inequality for upper bound holds with equality) for the convergence of REK and REGS. These bounds hold for all types of linear systems (consistent or inconsistent, overdetermined or underdetermined, $\mathbf{A}$ has full column rank or not) and are better than the existing ones.

Organization of this paper. In the rest of this section, we give some notation and preliminaries. In Section 2, we review the RK algorithm and the REK algorithm. We present a slightly different variant of REK and prove its convergence. In Section 3, we review the RGS algorithm and the REGS algorithm. We present a mathematically equivalent variant of REGS and prove its convergence. Numerical examples are given in Section 4 to illustrate the theoretical results. We present brief concluding remarks in Section 5.

TABLE 1 Summary of the convergence of RK, RGS, REK, and REGS in expectation to the Moore–Penrose pseudoinverse solution $\mathbf{A}^\dagger\mathbf{b}$ for all types of linear systems

Linear system	rank(A)	RK	RGS	REK	REGS
Consistent	$= n$	Y[1]	Y[2]	Y[5]	Y[11]
Consistent	$< n$	Y[5]	N[11]	Y[5]	Y[11]
Inconsistent	$= n$	N[3]	Y[2]	Y[5]	Y[11]
Inconsistent	$< n$	N[5]	N[11]	Y[5]	Y[17]

Note. Y means the algorithm is convergent and N means not. RK = randomized Kaczmarz; RGS = randomized Gauss–Seidel; REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel.

Notation and preliminaries. For any random variable ξ, let $\mathbb{E}[\xi]$ denote its expectation. For an integer $m \geq 1$, let $[m] := \{1, 2, 3, \ldots, m\}$. Throughout the paper, all vectors are assumed to be column vectors. For any vector $\mathbf{u} \in \mathbb{R}^m$, we use $\mathbf{u}^{\mathrm{T}}$, u_i, and $\|\mathbf{u}\|_2$ to denote the transpose, the ith entry, and the Euclidean norm of $\mathbf{u}$, respectively. We use $\mathbf{e}_j$ to denote the jth column of the identity matrix $\mathbf{I}$ whose order is clear from the context. For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, we use $\mathbf{A}^{\mathrm{T}}$, $\|\mathbf{A}\|_{\mathrm{F}}$, rank($\mathbf{A}$), range($\mathbf{A}$), null($\mathbf{A}$), $\sigma_1(\mathbf{A})$, and $\sigma_r(\mathbf{A})$ to denote the transpose, the Frobenius norm, the rank, the column space, the null space, the largest singular value, and the smallest nonzero singular value of $\mathbf{A}$, respectively. We denote the columns and rows of $\mathbf{A}$ by $\{\mathbf{a}_j\}_{j=1}^n$ and $\{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\}_{i=1}^m$, respectively. That is to say,

$$\mathbf{A} = [\mathbf{a}_1 \ \ \mathbf{a}_2 \ \ \ldots \ \ \mathbf{a}_n], \quad \mathbf{A}^{\mathrm{T}} = [\widetilde{\mathbf{a}}_1 \ \ \widetilde{\mathbf{a}}_2 \ \ \ldots \ \ \widetilde{\mathbf{a}}_m].$$

All the convergence results depend on the positive number ρ defined as

$$\rho := 1 - \frac{\sigma_r^2(\mathbf{A})}{\|\mathbf{A}\|_{\mathrm{F}}^2}.$$

The following lemmas will be used extensively in this paper. Their proofs are straightforward.

Lemma 1. *Let* $\mathbf{A}$ *be any nonzero real matrix. For every* $\mathbf{u} \in$ *range*($\mathbf{A}$)*, it holds*

$$\mathbf{u}^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{A}\mathbf{A}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)\mathbf{u} \leq \rho\|\mathbf{u}\|_2^2.$$

The equality holds if $\sigma_1(\mathbf{A}) = \sigma_r(\mathbf{A})$*, that is, all the nonzero singular values of* $\mathbf{A}$ *are the same.*

Lemma 2. *Let* $\mathbf{a}$ *be any nonzero real vector. Then,*

$$\left(\frac{\mathbf{a}\mathbf{a}^{\mathrm{T}}}{\|\mathbf{a}\|_2^2}\right)^2 = \frac{\mathbf{a}\mathbf{a}^{\mathrm{T}}}{\|\mathbf{a}\|_2^2}, \quad \left(\mathbf{I} - \frac{\mathbf{a}\mathbf{a}^{\mathrm{T}}}{\|\mathbf{a}\|_2^2}\right)^2 = \mathbf{I} - \frac{\mathbf{a}\mathbf{a}^{\mathrm{T}}}{\|\mathbf{a}\|_2^2}.$$

2 | RK AND ITS EXTENSION

Strohmer et al.[1] proposed the following RK algorithm (see Algorithm 1), where iterates are constructed by projections onto the hyperplanes defined by equations.

Algorithm 1 Randomized Kaczmarz [1] for **Ax=b**

Initialize $\mathbf{x}^0 \in \mathbb{R}^n$
for $k = 1, 2, \ldots$ **do**
  Pick $i \in [m]$ with probability $\|\widetilde{\mathbf{a}}_i\|_2^2 / \|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x}^{k-1} - b_i}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i$

If $\mathbf{Ax} = \mathbf{b}$ is consistent, in Theorem 3.4 of the work of Zouzias et al.,[5] it was proved that RK with initial guess $\mathbf{x}^0 \in \text{range}(\mathbf{A}^\mathrm{T})$ generates $\mathbf{x}^k$, which converges linearly in expectation to the Moore–Penrose pseudoinverse solution $\mathbf{A}^\dagger\mathbf{b}$:

$$\mathbb{E}\left[\left\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\right\|_2^2\right] \le \rho^k \left\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\right\|_2^2.$$

By the same approach as used in the proof of Theorem 3.2 in the work of Zouzias et al.,[5] we can prove the following theorem, which will be used to prove the tight upper bound for the convergence of REK.

Theorem 1. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$. *Let* $\mathbf{z}^k$ *denote the kth iterate of RK applied to* $\mathbf{A}^\mathrm{T}\mathbf{z} = \mathbf{0}$ *with initial guess* $\mathbf{z}^0 \in \mathbf{b} + \textit{range}(\mathbf{A})$. *In exact arithmetic, it holds*

$$\mathbb{E}\left[\left\|\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2\right] \le \rho^k \left\|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2.$$

Proof. The iteration is

$$\mathbf{z}^k = \mathbf{z}^{k-1} - \frac{\mathbf{a}_j^\mathrm{T}\mathbf{z}^{k-1}}{\|\mathbf{a}_j\|_2^2}\mathbf{a}_j.$$

By $\mathbf{a}_j^\mathrm{T}(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} = 0$ (because $\mathbf{A}^\mathrm{T}(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} = 0$), we have

$$\begin{aligned}
\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} &= \mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} - \frac{\mathbf{a}_j^\mathrm{T}\mathbf{z}^{k-1} - \mathbf{a}_j^\mathrm{T}(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}}{\|\mathbf{a}_j\|_2^2}\mathbf{a}_j \\
&= \mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} - \frac{\mathbf{a}_j^\mathrm{T}\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)}{\|\mathbf{a}_j\|_2^2}\mathbf{a}_j \\
&= \left(\mathbf{I} - \frac{\mathbf{a}_j\mathbf{a}_j^\mathrm{T}}{\|\mathbf{a}_j\|_2^2}\right)\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right).
\end{aligned}$$

By $\mathbf{z}^0 \in \mathbf{b} + \text{range}(\mathbf{A})$ and $\mathbf{A}\mathbf{A}^\dagger\mathbf{b} \in \text{range}(\mathbf{A})$, we have $\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} \in \text{range}(\mathbf{A})$. Then, it is easy to show that $\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b} \in \text{range}(\mathbf{A})$ by induction. Let $\mathbb{E}_{k-1}[\cdot]$ denote the conditional expectation conditioned on the first $k - 1$ iterations of RK. It follows that

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\left\|\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2\right] &= \mathbb{E}_{k-1}\left[\left(\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)^\mathrm{T}\left(\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)\right] \\
&= \mathbb{E}_{k-1}\left[\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)^\mathrm{T}\left(\mathbf{I} - \frac{\mathbf{a}_j\mathbf{a}_j^\mathrm{T}}{\|\mathbf{a}_j\|_2^2}\right)^2\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)\right] \\
&= \mathbb{E}_{k-1}\left[\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)^\mathrm{T}\left(\mathbf{I} - \frac{\mathbf{a}_j\mathbf{a}_j^\mathrm{T}}{\|\mathbf{a}_j\|_2^2}\right)\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)\right] \\
&= \left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right)^\mathrm{T}\left(\mathbf{I} - \frac{\mathbf{A}\mathbf{A}^\mathrm{T}}{\|\mathbf{A}\|_\mathrm{F}^2}\right)\left(\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right) \\
&\le \rho\left\|\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2 \quad \text{(by Lemma 1).}
\end{aligned}$$

Taking expectation gives

$$\mathbb{E}\left[\left\|\mathbf{z}^k - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2\right] \le \rho\mathbb{E}\left[\left\|\mathbf{z}^{k-1} - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2\right].$$

Unrolling the recurrence yields the result. □

If $\mathbf{Ax} = \mathbf{b}$ is inconsistent, Needell[3] and Zouzias et al.[5] showed that RK does not converge to $\mathbf{A}^\dagger\mathbf{b}$. To resolve this problem, Zouzias et al.[5] proposed the following REK algorithm (here, we call it REK-ZF; see Algorithm 2) by using Popa's extended Kacamarz method.[28–30] Zouzias et al.[5] used the starting vectors $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{z}^0 = \mathbf{b}$, and proved the convergence bound

$$\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \le \rho^{\lfloor k/2 \rfloor}\left(1 + \frac{2\sigma_1^2(\mathbf{A})}{\sigma_r^2(\mathbf{A})}\right)\|\mathbf{A}^\dagger\mathbf{b}\|_2^2. \tag{1}$$

Algorithm 2 REK-ZF [5]

Initialize $\mathbf{x}^0 \in \text{range}(\mathbf{A}^{\mathrm{T}})$ and $\mathbf{z}^0 \in \mathbf{b} + \text{range}(\mathbf{A})$
for $k = 1, 2, \ldots$ **do**
  Pick $j \in [n]$ with probability $\|\mathbf{a}_j\|_2^2/\|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{z}^k = \mathbf{z}^{k-1} - \frac{\mathbf{a}_j^{\mathrm{T}}\mathbf{z}^{k-1}}{\|\mathbf{a}_j\|_2^2}\mathbf{a}_j$
  Pick $i \in [m]$ with probability $\|\widetilde{\mathbf{a}}_i\|_2^2/\|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x}^{k-1} - b_i + z_i^{k-1}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i$

It was mentioned in the works of Liu et al.[7,14] that REK-ZF uses RK twice at each iteration: $\mathbf{z}^k$ is the kth iterate of RK applied to $\mathbf{A}^{\mathrm{T}}\mathbf{z} = \mathbf{0}$ with initial guess $\mathbf{z}^0$, and $\mathbf{x}^k$ is a one-step RK update for the linear system $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^{k-1}$ from $\mathbf{x}^{k-1}$. We note that REK-ZF updates the kth iterate $\mathbf{x}^k$ using the vector $\mathbf{z}^{k-1}$, although the vector $\mathbf{z}^k$ is already computed. Because the vector $\mathbf{z}^k$ is a better approximation to $(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}$ by Theorem 1, we present a slightly different randomized extended Kaczmarz algorithm (we call it REK-S; see Algorithm 3), which generates $\mathbf{x}^k$ by the one-step RK update for the linear system $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^k$ (used in REK-S) instead of $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^{k-1}$ (used in REK-ZF) from $\mathbf{x}^{k-1}$. By refining the quantities in the original proof for the convergence of REK,[5] we present a better convergence bound for REK-S in the following theorem.

Algorithm 3 REK-S

Initialize $\mathbf{x}^0 \in \text{range}(\mathbf{A}^{\mathrm{T}})$ and $\mathbf{z}^0 \in \mathbf{b} + \text{range}(\mathbf{A})$
for $k = 1, 2, \ldots$ **do**
  Pick $j \in [n]$ with probability $\|\mathbf{a}_j\|_2^2/\|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{z}^k = \mathbf{z}^{k-1} - \frac{\mathbf{a}_j^{\mathrm{T}}\mathbf{z}^{k-1}}{\|\mathbf{a}_j\|_2^2}\mathbf{a}_j$
  Pick $i \in [m]$ with probability $\|\widetilde{\mathbf{a}}_i\|_2^2/\|\mathbf{A}\|_{\mathrm{F}}^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x}^{k-1} - b_i + z_i^{k}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i$

Theorem 2. *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{x}^k$ denote the kth iterate of REK-S with $\mathbf{x}^0 \in$ range$(\mathbf{A}^{\mathrm{T}})$ and $\mathbf{z}^0 \in \mathbf{b}$ + range$(\mathbf{A})$. In exact arithmetic, it holds*

$$\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \le \frac{k\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\left\|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2 + \rho^k\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2. \tag{2}$$

Proof. Let

$$\hat{\mathbf{x}}^k = \mathbf{x}^{k-1} - \frac{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x}^{k-1} - b_i + \mathbf{e}_i^{\mathrm{T}}(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i,$$

which is actually the one-step RK update for the linear system $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^\dagger\mathbf{b}$ from $\mathbf{x}^{k-1}$. We have

$$\begin{aligned}
\hat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b} &= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x}^{k-1} - \mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{A}^\dagger\mathbf{b}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i \\
&= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{x}^{k-1} - \widetilde{\mathbf{a}}_i^{\mathrm{T}}\mathbf{A}^\dagger\mathbf{b}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i \\
&= \left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b})
\end{aligned}$$

and

$$\mathbf{x}^k - \widehat{\mathbf{x}}^k = \frac{\mathbf{e}_i^{\mathrm{T}}\left((\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}-\mathbf{z}^k\right)}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i.$$

By the orthogonality $(\widehat{\mathbf{x}}^k - \mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}(\mathbf{x}^k - \widehat{\mathbf{x}}^k) = 0$ (which is obvious from the above two equations), we have

$$\|\mathbf{x}^k - \mathbf{A}^{\dagger}\mathbf{b}\|_2^2 = \|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2 + \|\widehat{\mathbf{x}}^k - \mathbf{A}^{\dagger}\mathbf{b}\|_2^2. \tag{3}$$

Let $\mathbb{E}_{k-1}[\cdot]$ denote the conditional expectation conditioned on the first $k-1$ iterations of REK-S. That is,

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

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

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

Similarly, we denote the conditional expectation conditioned on the first $k-1$ iterations and the kth row chosen as

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

Then, by the law of total expectation, we have

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

It follows from

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right] &= \mathbb{E}_{k-1}\left[\frac{\left(\mathbf{e}_i^{\mathrm{T}}\left((\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}-\mathbf{z}^k\right)\right)^2}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right] \\
&= \mathbb{E}_{k-1}^{j}\left[\mathbb{E}_{k-1}^{i}\left[\frac{\left(\mathbf{e}_i^{\mathrm{T}}\left((\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}-\mathbf{z}^k\right)\right)^2}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right]\right] \\
&= \mathbb{E}_{k-1}^{j}\left[\frac{\left\|\mathbf{z}^k - (\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}\right\|_2^2}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right] \\
&= \frac{1}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}_{k-1}\left[\left\|\mathbf{z}^k - (\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}\right\|_2^2\right]
\end{aligned}$$

that

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{x}^k - \widehat{\mathbf{x}}^k\|_2^2\right] &= \frac{1}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}\left[\left\|\mathbf{z}^k - (\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}\right\|_2^2\right] \\
&\leq \frac{\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\left\|\mathbf{z}^0 - (\mathbf{I}-\mathbf{A}\mathbf{A}^{\dagger})\mathbf{b}\right\|_2^2 \quad \text{(by Theorem 1)}.
\end{aligned} \tag{4}$$

By $\mathbf{x}^0 \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$ and $\mathbf{A}^{\dagger}\mathbf{b} \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$, we have $\mathbf{x}^0 - \mathbf{A}^{\dagger}\mathbf{b} \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$. Then, it is easy to show that $\mathbf{x}^k - \mathbf{A}^{\dagger}\mathbf{b} \in \mathrm{range}(\mathbf{A}^{\mathrm{T}})$ by induction. It follows from

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\|\widehat{\mathbf{x}}^k - \mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right] &= \mathbb{E}_{k-1}\left[(\widehat{\mathbf{x}}^k - \mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}(\widehat{\mathbf{x}}^k - \mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= \mathbb{E}_{k-1}\left[(\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)^2(\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= \mathbb{E}_{k-1}\left[(\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= (\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)(\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b}) \\
&\leq \rho\|\mathbf{x}^{k-1} - \mathbf{A}^{\dagger}\mathbf{b}\|_2^2 \quad \text{(by Lemma 1)}
\end{aligned}$$

that

$$\mathbb{E}\left[\|\hat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \le \rho\mathbb{E}\left[\|\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right]. \tag{5}$$

Combining (3), (4), and (5) yields

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] &= \mathbb{E}\left[\|\mathbf{x}^k - \hat{\mathbf{x}}^k\|_2^2\right] + \mathbb{E}\left[\|\hat{\mathbf{x}}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\le \frac{\rho^k}{\|\mathbf{A}\|_F^2}\left\|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2 + \rho\mathbb{E}\left[\|\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\le \frac{2\rho^k}{\|\mathbf{A}\|_F^2}\left\|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2 + \rho^2\mathbb{E}\left[\|\mathbf{x}^{k-2} - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \\
&\le \dots \le \frac{k\rho^k}{\|\mathbf{A}\|_F^2}\left\|\mathbf{z}^0 - (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{b}\right\|_2^2 + \rho^k\|\mathbf{x}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2.
\end{aligned}$$

This completes the proof. □

Remark 1. Substituting $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{z}^0 = \mathbf{b}$ into the convergence bound (2) yields

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] &\le \rho^k\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 + \frac{k\rho^k}{\|\mathbf{A}\|_F^2}\|\mathbf{A}\mathbf{A}^\dagger\mathbf{b}\|_2^2 \\
&\le \rho^k\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 + \frac{k\rho^k\sigma_1^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\left\|\mathbf{A}^\dagger\mathbf{b}\right\|_2^2 \\
&< \rho^{\lfloor k/2\rfloor}\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 + 2\lceil k/2\rceil\rho^k\frac{\sigma_1^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 \\
&= \rho^{\lfloor k/2\rfloor}\left(1 + 2\lceil k/2\rceil\rho^{k-\lfloor k/2\rfloor}\frac{\sigma_1^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\right)\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 \\
&< \rho^{\lfloor k/2\rfloor}\left(1 + 2\frac{\sigma_1^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\sum_{l=0}^{\infty}\rho^l\right)\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 \\
&= \rho^{\lfloor k/2\rfloor}\left(1 + 2\frac{\sigma_1^2(\mathbf{A})}{\|\mathbf{A}\|_F^2}\frac{1}{1-\rho}\right)\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 \\
&= \rho^{\lfloor k/2\rfloor}\left(1 + \frac{2\sigma_1^2(\mathbf{A})}{\sigma_r^2(\mathbf{A})}\right)\|\mathbf{A}^\dagger\mathbf{b}\|_2^2.
\end{aligned}$$

Hence, the convergence rate in our bound (2) is better, uniformly with respect to the iteration index, than that in the existing bound (1). Obviously, there is no matrix such that the bound (1) holds with equality, which means that the bound (1) is not tight. By Lemma 1, if $\sigma_1(\mathbf{A}) = \sigma_r(\mathbf{A})$, then all the inequalities in Theorems 1 and 2 become equalities, which means that our upper bound (2) is tight.

3 | RGS AND ITS EXTENSION

Leventhal et al.[2] proposed the following RGS algorithm (Algorithm 4; also called the randomized coordinate descent algorithm). The following theorem is a restatement of Lemma 4.2 in the work of Ma et al.[11] and will be used to prove the tight upper bound for REGS. Here, we provide a proof for completeness.

Algorithm 4 Randomized Gauss–Seidel[2]

Initialize $\mathbf{x}^0 \in \mathbb{R}^n$
for $k = 1, 2, \dots$ **do**
  Pick $j \in [n]$ with probability $\|\mathbf{a}_j\|_2^2/\|\mathbf{A}\|_F^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{a}_j^T(\mathbf{A}\mathbf{x}^{k-1}-\mathbf{b})}{\|\mathbf{a}_j\|_2^2}\mathbf{e}_j$

Theorem 3. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$. *Let* $\mathbf{x}^k$ *denote the kth iterate of RGS applied to* $\mathbf{Ax} = \mathbf{b}$ *with arbitrary* $\mathbf{x}^0 \in \mathbb{R}^n$. *In exact arithmetic, it holds*

$$\mathbb{E}\left[\|\mathbf{Ax}^k - \mathbf{AA}^\dagger\mathbf{b}\|_2^2\right] \le \rho^k \left\|\mathbf{Ax}^0 - \mathbf{AA}^\dagger\mathbf{b}\right\|_2^2.$$

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

$$\begin{aligned}
\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b} &= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\mathbf{a}_j^{\mathrm{T}}(\mathbf{Ax}^{k-1} - \mathbf{b})}{\|\mathbf{a}_j\|_2^2}\mathbf{e}_j \\
&= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\mathbf{e}_j^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{Ax}^{k-1} - \mathbf{A}^{\mathrm{T}}\mathbf{b})}{\|\mathbf{a}_j\|_2^2}\mathbf{e}_j \\
&= \mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b} - \frac{\mathbf{e}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b})}{\|\mathbf{a}_j\|_2^2}\mathbf{e}_j \\
&= \left(\mathbf{I} - \frac{\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{a}_j\|_2^2}\right)(\mathbf{x}^{k-1} - \mathbf{A}^\dagger\mathbf{b}),
\end{aligned}$$

which yields

$$\mathbf{Ax}^k - \mathbf{AA}^\dagger\mathbf{b} = \left(\mathbf{I} - \frac{\mathbf{a}_j\mathbf{a}_j^{\mathrm{T}}}{\|\mathbf{a}_j\|_2^2}\right)(\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b}).$$

It follows that

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\|\mathbf{Ax}^k - \mathbf{AA}^\dagger\mathbf{b}\|_2^2\right] &= \mathbb{E}_{k-1}\left[(\mathbf{Ax}^k - \mathbf{AA}^\dagger\mathbf{b})^{\mathrm{T}}(\mathbf{Ax}^k - \mathbf{AA}^\dagger\mathbf{b})\right] \\
&= \mathbb{E}_{k-1}\left[(\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b})^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{a}_j\mathbf{a}_j^{\mathrm{T}}}{\|\mathbf{a}_j\|_2^2}\right)^2(\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b})\right] \\
&= \mathbb{E}_{k-1}\left[(\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b})^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{a}_j\mathbf{a}_j^{\mathrm{T}}}{\|\mathbf{a}_j\|_2^2}\right)(\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b})\right] \\
&= (\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b})^{\mathrm{T}}\left(\mathbf{I} - \frac{\mathbf{AA}^{\mathrm{T}}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)(\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b}) \\
&\le \rho\|\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b}\|_2^2 \quad \text{(by Lemma 1)}.
\end{aligned}$$

Taking expectation gives

$$\mathbb{E}\left[\|\mathbf{Ax}^k - \mathbf{AA}^\dagger\mathbf{b}\|_2^2\right] \le \rho\mathbb{E}\left[\|\mathbf{Ax}^{k-1} - \mathbf{AA}^\dagger\mathbf{b}\|_2^2\right].$$

Unrolling the recurrence yields the result. □

If $\mathbf{A}$ has full column rank, Theorem 3 implies that $\mathbf{x}^k$ converges linearly in expectation to $\mathbf{A}^\dagger\mathbf{b}$. If $\mathbf{A}$ does not have full column rank, RGS fails to converge (see section 3.3 of the work of Ma et al.[11]). Ma et al.[11] proposed the following REGS algorithm (we call it REGS-MNR; see Algorithm 5) to resolve this problem. They used the starting vectors $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{z}^0 = \mathbf{0}$, and proved the convergence bound (see theorem 4.1 of the work of Ma et al.[11] for details)

$$\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \le \rho^k\|\mathbf{A}^\dagger\mathbf{b}\|_2^2 + \frac{2\rho^{\lfloor k/2\rfloor}}{\sigma_r^2(\mathbf{A})}\|\mathbf{AA}^\dagger\mathbf{b}\|_2^2. \tag{6}$$

Next, we study the convergence of REGS for a general linear system (consistent or inconsistent, full rank or rank deficient). We emphasize that the convergence analysis is similar as that of REK. To get a similar bound as that in Theorem 2 and for the convenience of discussion, we present the following REGS algorithm (we call it REGS-E; see Algorithm 6), which is mathematically equivalent to REGS-MNR. Actually, in exact arithmetic, the vector $\mathbf{z}^k$ in REGS-E is equal to the vector $\mathbf{x}^k - \mathbf{z}^k$ in REGS-MNR. We note that, at each iteration of REGS-E, $\mathbf{x}^k$ is the kth iterate of RGS and $\mathbf{z}^k$ is the one-step

RK update for the linear system $\mathbf{Az} = \mathbf{Ax}^k$ from $\mathbf{z}^{k-1}$. In the following theorem, we show that the vector $\mathbf{z}^k$ in REGS-E converges linearly in expectation to $\mathbf{A}^\dagger\mathbf{b}$. Our proof is almost the same as that in theorem 4.1 of the work of Ma et al.[11] but refines the quantities involved in the inequalities.

Algorithm 5 REGS-MNR[11]

Initialize $\mathbf{x}^0 \in \mathbb{R}^n$ and $\mathbf{z}^0 \in \mathbf{x}^0 + \text{range}(\mathbf{A}^\mathrm{T})$
for $k = 1, 2, \ldots$ **do**
  Pick $j \in [n]$ with probability $\|\mathbf{a}_j\|_2^2/\|\mathbf{A}\|_\mathrm{F}^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{a}_j^\mathrm{T}(\mathbf{Ax}^{k-1}-\mathbf{b})}{\|\mathbf{a}_j\|_2^2}\mathbf{e}_j$
  Pick $i \in [m]$ with probability $\|\widetilde{\mathbf{a}}_i\|_2^2/\|\mathbf{A}\|_\mathrm{F}^2$
  Set $\mathbf{P}_i = \mathbf{I} - \widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}/\|\widetilde{\mathbf{a}}_i\|_2^2$
  Set $\mathbf{z}^k = \mathbf{P}_i(\mathbf{z}^{k-1} + \mathbf{x}^k - \mathbf{x}^{k-1})$
Output $\mathbf{x}^t - \mathbf{z}^t$ at some step t as the estimated solution

Algorithm 6 REGS-E

Initialize $\mathbf{x}^0 \in \mathbb{R}^n$ and $\mathbf{z}^0 \in \text{range}(\mathbf{A}^\mathrm{T})$
for $k = 1, 2, \ldots$ **do**
  Pick $j \in [n]$ with probability $\|\mathbf{a}_j\|_2^2/\|\mathbf{A}\|_\mathrm{F}^2$
  Set $\mathbf{x}^k = \mathbf{x}^{k-1} - \frac{\mathbf{a}_j^\mathrm{T}(\mathbf{Ax}^{k-1}-\mathbf{b})}{\|\mathbf{a}_j\|_2^2}\mathbf{e}_j$
  Pick $i \in [m]$ with probability $\|\widetilde{\mathbf{a}}_i\|_2^2/\|\mathbf{A}\|_\mathrm{F}^2$
  Set $\mathbf{z}^k = \mathbf{z}^{k-1} - \frac{\widetilde{\mathbf{a}}_i^\mathrm{T}(\mathbf{z}^{k-1}-\mathbf{x}^k)}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i$

Theorem 4. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$. *Let* $\mathbf{z}^k$ *denote the kth iterate of REGS-E with arbitrary* $\mathbf{x}^0 \in \mathbb{R}^n$ *and* $\mathbf{z}^0 \in$ *range*$(\mathbf{A}^\mathrm{T})$. *In exact arithmetic, it holds*

$$\mathbb{E}\left[\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2\right] \le \rho^k\|\mathbf{z}^0 - \mathbf{A}^\dagger\mathbf{b}\|_2^2 + \frac{k\rho^k}{\|\mathbf{A}\|_\mathrm{F}^2}\|\mathbf{Ax}^0 - \mathbf{AA}^\dagger\mathbf{b}\|_2^2. \tag{7}$$

Proof. By $\mathbf{z}^0 \in \text{range}(\mathbf{A}^\mathrm{T})$ and $\mathbf{A}^\dagger\mathbf{b} \in \text{range}(\mathbf{A}^\mathrm{T})$, we have $\mathbf{z}^0 - \mathbf{A}^\dagger\mathbf{b} \in \text{range}(\mathbf{A}^\mathrm{T})$. Then, it is easy to show that $\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b} \in \text{range}(\mathbf{A}^\mathrm{T})$ by induction. We now analyze the norm of $\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}$. Note that

$$\begin{aligned}
\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b} &= \mathbf{z}^{k-1} - \frac{\widetilde{\mathbf{a}}_i^\mathrm{T}(\mathbf{z}^{k-1} - \mathbf{x}^k)}{\|\widetilde{\mathbf{a}}_i\|_2^2}\widetilde{\mathbf{a}}_i - \mathbf{A}^\dagger\mathbf{b} \\
&= \left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)\mathbf{z}^{k-1} + \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b} \\
&= \left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1} - \mathbf{A}^\dagger\mathbf{b}) + \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}(\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}).
\end{aligned}$$

It follows from the orthogonality, namely,

$$(\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b})^\mathrm{T}\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1} - \mathbf{A}^\dagger\mathbf{b}) = 0,$$

that

$$\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2 = \left\|\left(\mathbf{I} - \frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^\mathrm{T}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1} - \mathbf{A}^\dagger\mathbf{b})\right\|_2^2 + \left\|\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}\,2}}{\|\widetilde{\mathbf{a}}_i\|_2}(\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b})\right\|_2^2. \tag{8}$$

It follows from

$$
\begin{aligned}
\mathbb{E}_{k-1}\left[\left\|\left(\mathbf{I}-\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})\right\|_2^2\right] &= \mathbb{E}_{k-1}\left[(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\mathbf{I}-\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)^2(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= \mathbb{E}_{k-1}\left[(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\mathbf{I}-\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= (\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\mathbf{I}-\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\right)(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b}) \\
&\le \rho\|\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b}\|_2^2 \quad \text{(by Lemma 1)}
\end{aligned}
$$

that

$$
\mathbb{E}\left[\left\|\left(\mathbf{I}-\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})\right\|_2^2\right] \le \rho\mathbb{E}\left[\|\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right]. \tag{9}
$$

It follows from

$$
\begin{aligned}
\mathbb{E}_{k-1}\left[\left\|\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})\right\|_2^2\right] &= \mathbb{E}_{k-1}\left[(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\left(\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)^2(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= \mathbb{E}_{k-1}^{j}\left[\mathbb{E}_{k-1}^{i}\left[(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})\right]\right] \\
&= \mathbb{E}_{k-1}^{j}\left[(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})^{\mathrm{T}}\frac{\mathbf{A}^{\mathrm{T}}\mathbf{A}}{\|\mathbf{A}\|_{\mathrm{F}}^2}(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})\right] \\
&= \frac{1}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}_{k-1}\left[\|\mathbf{A}\mathbf{x}^{k}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right]
\end{aligned}
$$

that

$$
\begin{aligned}
\mathbb{E}\left[\left\|\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})\right\|_2^2\right] &= \frac{1}{\|\mathbf{A}\|_{\mathrm{F}}^2}\mathbb{E}\left[\|\mathbf{A}\mathbf{x}^{k}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right] \\
&\le \frac{\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{x}^{0}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\|_2^2 \quad \text{(by Theorem 3).}
\end{aligned} \tag{10}
$$

Combining (8), (9), and (10) yields

$$
\begin{aligned}
\mathbb{E}\left[\|\mathbf{z}^{k}-\mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right] &= \mathbb{E}\left[\left\|\left(\mathbf{I}-\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}\right)(\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b})\right\|_2^2\right] + \mathbb{E}\left[\left\|\frac{\widetilde{\mathbf{a}}_i\widetilde{\mathbf{a}}_i^{\mathrm{T}}}{\|\widetilde{\mathbf{a}}_i\|_2^2}(\mathbf{x}^{k}-\mathbf{A}^{\dagger}\mathbf{b})\right\|_2^2\right] \\
&\le \rho\mathbb{E}\left[\|\mathbf{z}^{k-1}-\mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right] + \frac{\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{x}^{0}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\|_2^2 \\
&\le \rho^2\mathbb{E}\left[\|\mathbf{z}^{k-2}-\mathbf{A}^{\dagger}\mathbf{b}\|_2^2\right] + \frac{2\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{x}^{0}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\|_2^2 \\
&\le \ \dots \ \le \rho^k\|\mathbf{z}^{0}-\mathbf{A}^{\dagger}\mathbf{b}\|_2^2 + \frac{k\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{x}^{0}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}\|_2^2.
\end{aligned}
$$

This completes the proof. □

Remark 2. Substituting $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{z}^0 = \mathbf{0}$ into the convergence bound (7) yields

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{z}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2\right] &\le \rho^k \|\mathbf{A}^\dagger \mathbf{b}\|_2^2 + \frac{k\rho^k}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{A}^\dagger \mathbf{b}\|_2^2 \\
&\le \rho^k \|\mathbf{A}^\dagger \mathbf{b}\|_2^2 + \frac{2\lceil k/2\rceil \rho^{k-\lfloor k/2\rfloor}\rho^{\lfloor k/2\rfloor}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{A}^\dagger \mathbf{b}\|_2^2 \\
&< \rho^k \|\mathbf{A}^\dagger \mathbf{b}\|_2^2 + \frac{2\rho^{\lfloor k/2\rfloor}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{A}^\dagger \mathbf{b}\|_2^2 \sum_{l=0}^{\infty}\rho^l \\
&= \rho^k \|\mathbf{A}^\dagger \mathbf{b}\|_2^2 + \frac{2\rho^{\lfloor k/2\rfloor}}{\|\mathbf{A}\|_{\mathrm{F}}^2}\|\mathbf{A}\mathbf{A}^\dagger \mathbf{b}\|_2^2 \frac{1}{1-\rho} \\
&= \rho^k \|\mathbf{A}^\dagger \mathbf{b}\|_2^2 + \frac{2\rho^{\lfloor k/2\rfloor}}{\sigma_r^2(\mathbf{A})}\|\mathbf{A}\mathbf{A}^\dagger \mathbf{b}\|_2^2.
\end{aligned}$$

Hence, the convergence rate in our bound (7) is better, uniformly with respect to the iteration index, than that in the existing bound (6). Obviously, there is no matrix such that the bound (6) holds with equality, which means that the bound (6) is not tight. By Lemma 1, if $\sigma_1(\mathbf{A}) = \sigma_r(\mathbf{A})$, then all the inequalities in Theorems 3 and 4 become equalities, which means that our upper bound (7) is tight.

TABLE 2 The detailed features of the six examples

Example	Linear system	Matrix	m	n	rank	$\sigma_1(\mathbf{A})/\sigma_r(\mathbf{A})$	ρ
1	consistent	$\mathbf{UDV}^{\mathrm{T}}$	500	250	250	1.25	0.9960
2	consistent	$\mathbf{UDV}^{\mathrm{T}}$	500	250	150	1.5	0.9957
3	inconsistent	$\mathbf{UDV}^{\mathrm{T}}$	500	250	250	1.75	0.9979
4	inconsistent	$\mathbf{UDV}^{\mathrm{T}}$	500	250	150	2	0.9971
5	consistent	`bibd_17_8`	136	24,310	136	9.0370	0.9975
6	inconsistent	`relat6`	2,340	157	137	7.7377	0.9996

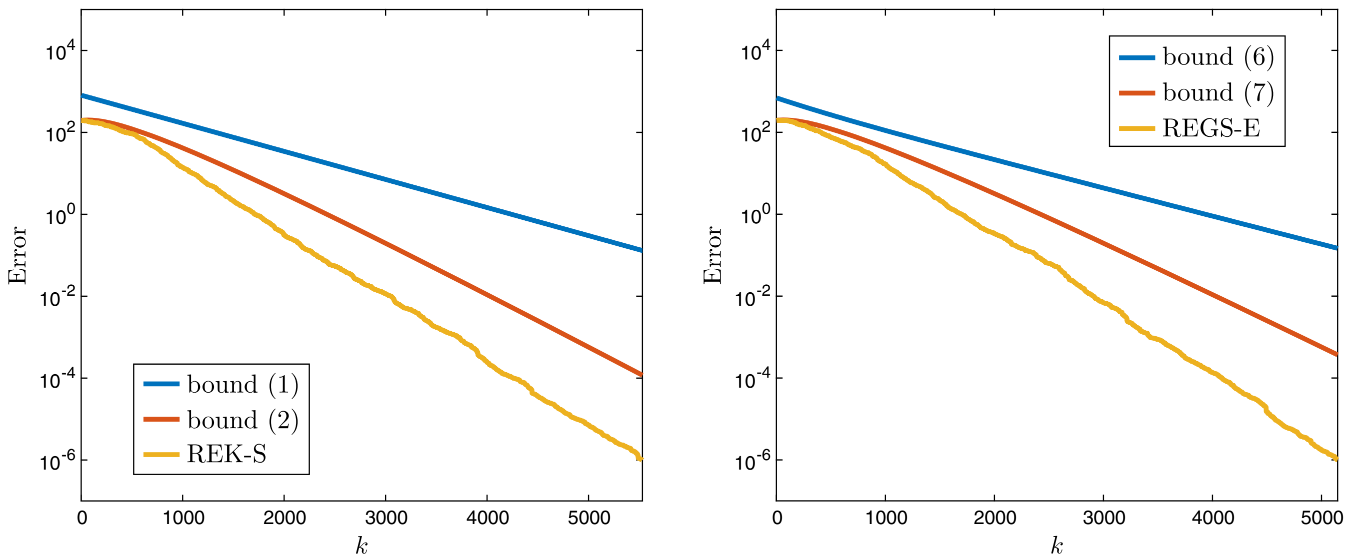

FIGURE 1 The error $\|\mathbf{x}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2$ (log-scale) for REK-S (left) and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger \mathbf{b}\|_2^2$ (log-scale) for REGS-E (right) on a consistent linear system with full column rank $\mathbf{A}$: $m = 500$, $n = 250$, $r = 250$, $\sigma_1(\mathbf{A}) = 1.25$, $\sigma_r(\mathbf{A}) = 1$, and $\rho = 0.9960$. REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel

4 | NUMERICAL RESULTS

In this section, we report the numerical results of the REK-S algorithm and the REGS-E algorithm for solving different types of linear systems. All experiments are performed using MATLAB on a laptop with 2.7-GHz Intel Core i7 processor, 16-GB memory, and Mac operating system.

We compare our bounds (2) and (7) with the existing bounds (1) and (6) via six examples. The coefficient matrices of the first four examples are generated as follows. For given m, n, $r = \mathrm{rank}(\mathbf{A})$, $\sigma_1(\mathbf{A})$, and $\sigma_r(\mathbf{A})$, 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. The matrix $\mathbf{D}$ is an $r \times r$ diagonal matrix whose first $r - 2$ diagonal entries are uniformly distributed numbers in $[\sigma_r(\mathbf{A}), \sigma_1(\mathbf{A})]$, and the last two diagonal entries are $\sigma_r(\mathbf{A})$ and $\sigma_1(\mathbf{A})$. The coefficient

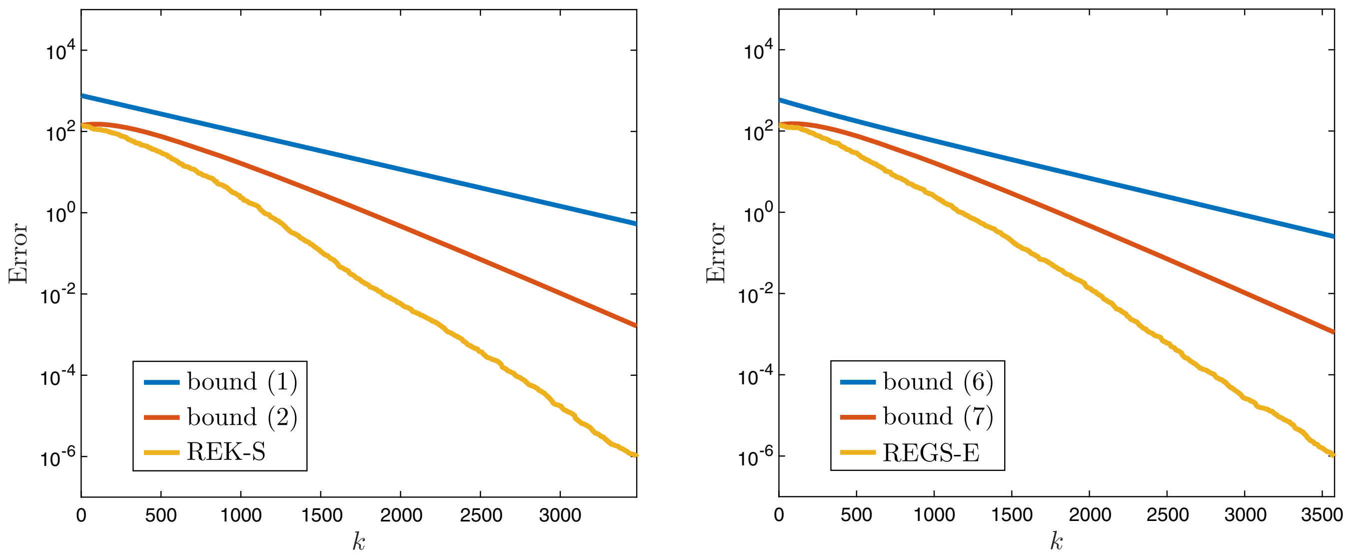

FIGURE 2 The error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REK-S (left) and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REGS-E (right) on a consistent linear system with rank-deficient $\mathbf{A}$: $m = 500$, $n = 250$, $r = 150$, $\sigma_1(\mathbf{A}) = 1.5$, $\sigma_r(\mathbf{A}) = 1$, and $\rho \approx 0.9957$. REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel

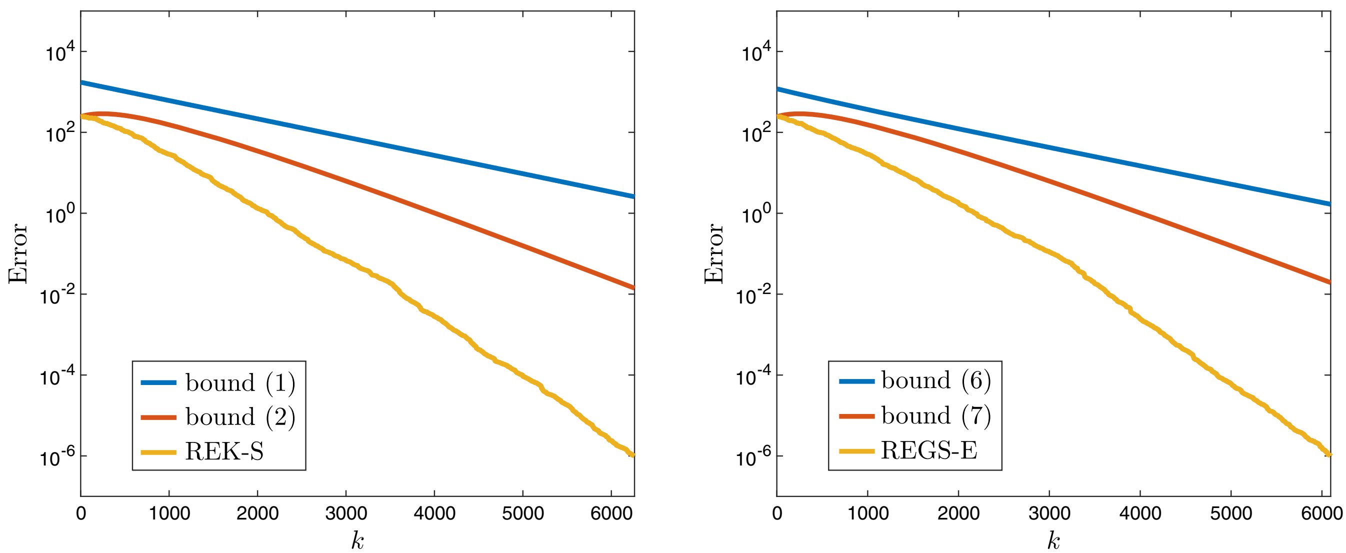

FIGURE 3 The error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REK-S (left) and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REGS-E (right) on an inconsistent linear system with full column rank $\mathbf{A}$: $m = 500$, $n = 250$, $r = 250$, $\sigma_1(\mathbf{A}) = 1.75$, $\sigma_r(\mathbf{A}) = 1$, and $\rho \approx 0.9979$. REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel

matrices of the last two examples are `bibd_17_8` and `relat6` from the University of Florida sparse matrix collection.[31] To construct a consistent linear system, we set $\mathbf{b} = \mathbf{Ax}$ where $\mathbf{x}$ is a vector with entries generated from a standard normal distribution. To construct an inconsistent linear system, we set $\mathbf{b} = \mathbf{Ax} + \mathbf{r}$ where $\mathbf{x}$ is a vector with entries generated from a standard normal distribution and the residual $\mathbf{r} \in \text{null}(\mathbf{A}^{\mathrm{T}})$. Note that one can obtain such a vector $\mathbf{r}$ by the MATLAB function `null`. All problem instances are listed in Table 2.

In Figures 1–6, we plot the error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ for REK-S with $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{z}^0 = \mathbf{b}$ and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ for REGS-E with $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{z}^0 = \mathbf{0}$ for the six examples, respectively. For all cases, REK-S and REGS-E converge to the Moore–Penrose pseudoinverse solution $\mathbf{A}^\dagger\mathbf{b}$ at about the same rate, and our bounds (2) and (7) are better than the existing bounds (1) and (6).

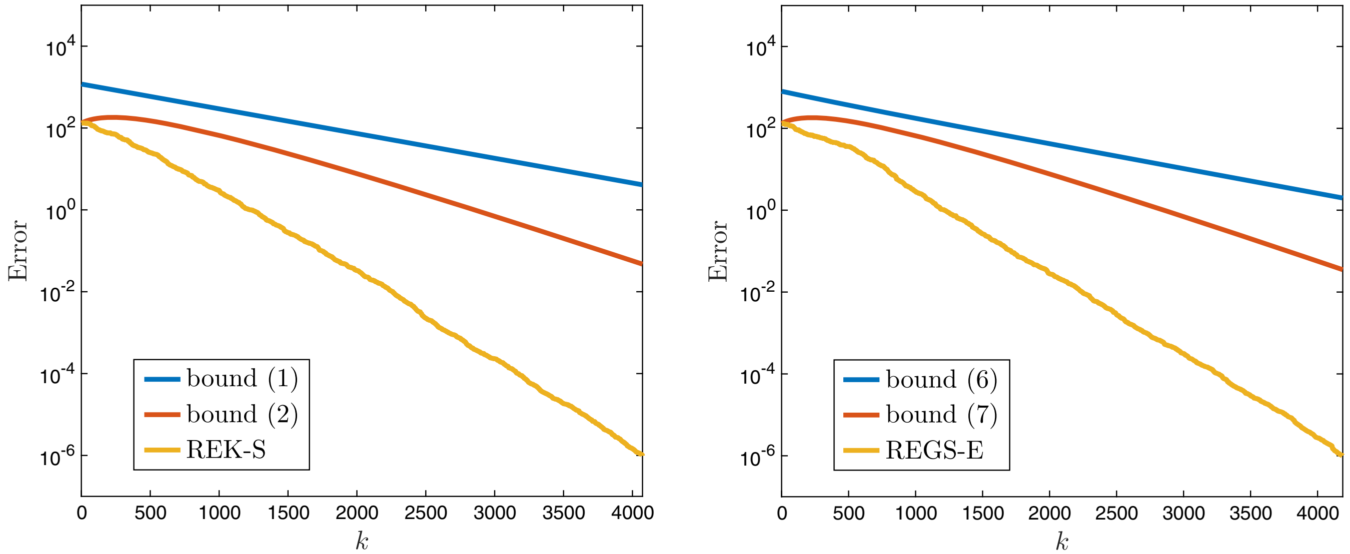

FIGURE 4 The error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REK-S (left) and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REGS-E (right) on an inconsistent linear system with rank-deficient $\mathbf{A}$: $m = 500$, $n = 250$, $r = 150$, $\sigma_1(\mathbf{A}) = 2$, $\sigma_r(\mathbf{A}) = 1$, and $\rho \approx 0.9971$. REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel

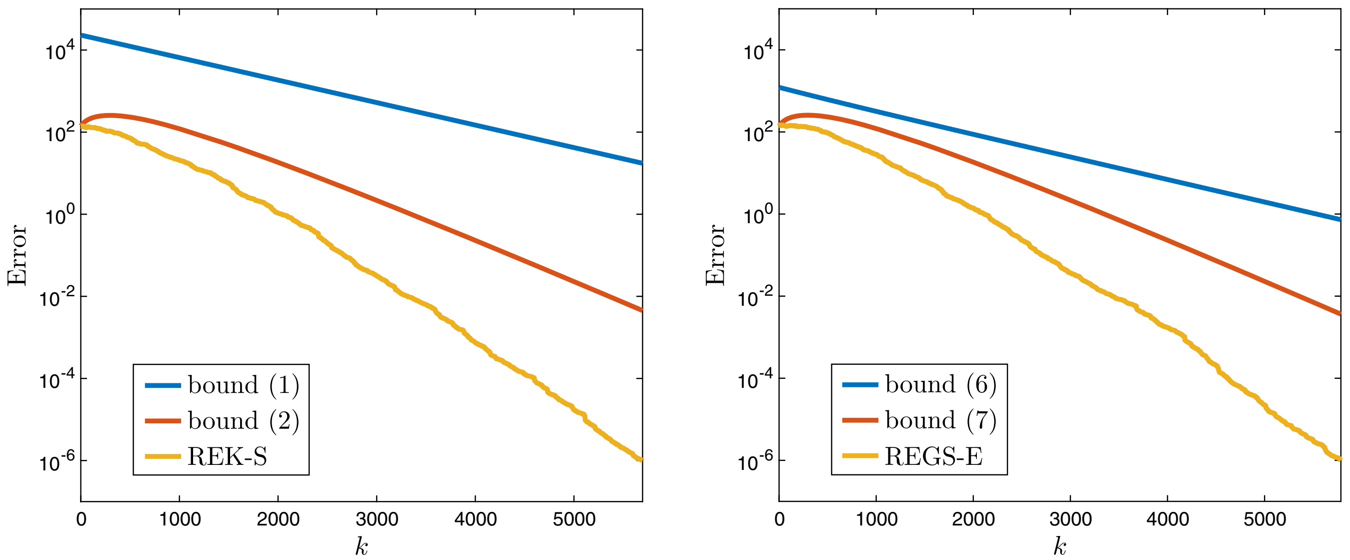

FIGURE 5 The error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REK-S (left) and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REGS-E (right) on a consistent linear system with the full row rank matrix `bibd_17_8`: $m = 136$, $n = 24310$, $r = 136$, $\sigma_1(\mathbf{A}) \approx 374.3528$, $\sigma_r(\mathbf{A}) \approx 41.4246$, and $\rho \approx 0.9975$. REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel

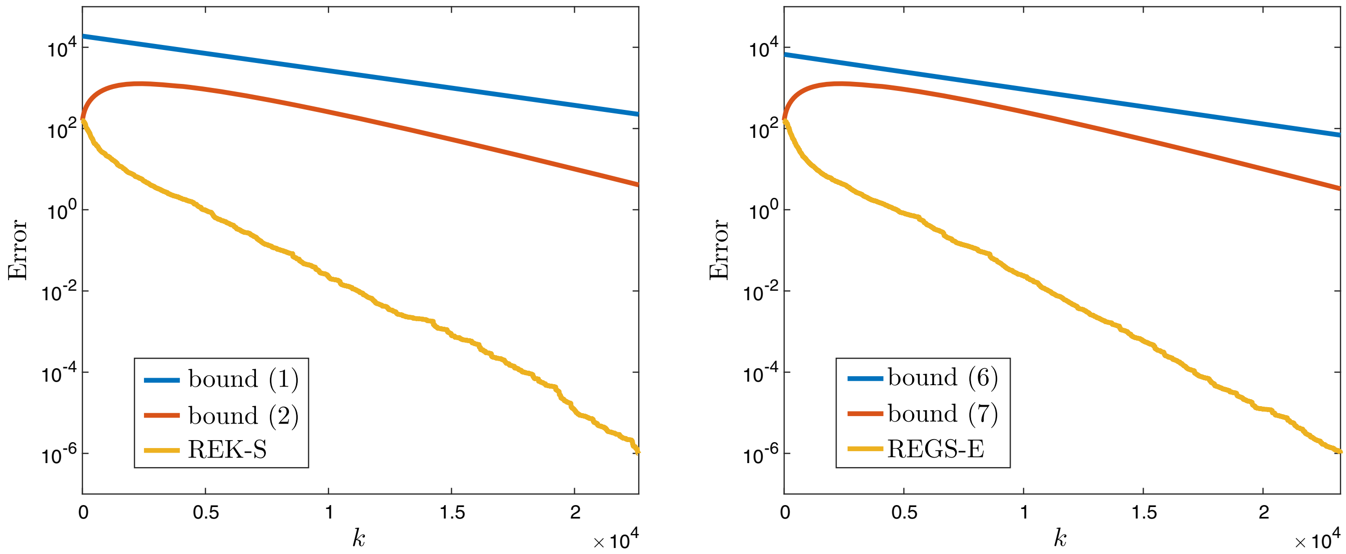

FIGURE 6 The error $\|\mathbf{x}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REK-S (left) and the error $\|\mathbf{z}^k - \mathbf{A}^\dagger\mathbf{b}\|_2^2$ (log-scale) for REGS-E (right) on an inconsistent linear system with the rank-deficient matrix `relat6`: $m = 2,340$, $n = 157$, $r = 137$, $\sigma_1(\mathbf{A}) \approx 14.1632$, $\sigma_r(\mathbf{A}) \approx 1.8304$, and $\rho \approx 0.9996$. REK = randomized extended Kaczmarz; REGS = randomized extended Gauss–Seidel

5 | CONCLUDING REMARKS

We have proposed tight upper bounds for the convergence of the REK and REGS algorithms. These upper bounds are attained for the case that all nonzero singular values of **A** are the same. Our convergence analysis applies to all types of linear systems. Numerical experiments confirm the theoretical results.

Recently, in the works of Zouzias et al.[32] and Loizou et al.,[33,34] it was shown that randomized Kaczmarz-type methods and RGS can work as randomized gossip algorithms for solving the average consensus problem. The proposed convergence analysis in this paper could be useful for the analysis of randomized gossip algorithms. The extension of the proposed methods and their analysis to their block variants where in each step blocks of rows and columns are used to compute the next iterate and the extension of the acceleration techniques[4,14,17] improving the performance of randomized algorithms to REK and REGS will be the future work.

ACKNOWLEDGEMENTS

The author would like to thank the referees for the detailed comments and suggestions that have led to many improvements. The research of the author was supported by the National Natural Science Foundation of China (11771364 and 91430213) and the Fundamental Research Funds for the Central Universities (20720160002). There are no conflicts of interest to this work.

ORCID

Kui Du https://orcid.org/0000-0003-4891-0081

REFERENCES

1. Strohmer T, Vershynin R. A randomized Kaczmarz algorithm with exponential convergence. J Fourier Anal Appl. 2009;15(2):262. https://doi.org/10.1007/s00041-008-9030-4
2. Leventhal D, Lewis AS. Randomized methods for linear constraints: convergence rates and conditioning. Math Oper Res. 2010;35(3):513–720. https://doi.org/10.1287/moor.1100.0456
3. Needell D. Randomized Kaczmarz solver for noisy linear systems. BIT Numer Math. 2010;50(2):395–403. https://doi.org/10.1007/s10543-010-0265-5
4. Eldar YC, Needell D. Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss Lemma. Numer Algorithm. 2011;58(2):163–177. https://doi.org/10.1007/s11075-011-9451-z
5. Zouzias A, Freris NM. Randomized extended Kaczmarz for solving least squares. SIAM J Matrix Anal Appl. 2013;34(2):773–793. https://doi.org/10.1137/120889897

6. Dai L, Schön TB. On the exponential convergence of the Kaczmarz algorithm. IEEE Signal Process Lett. 2015;22(10):1571–1574.
7. Liu J, Wright SJ, Sridhar S. An asynchronous parallel randomized Kaczmarz algorithm. 2014 [cited 2014 Jun 7]. Available from: https://arxiv.org/pdf/1401.4780
8. Needell D, Tropp JA. Paved with good intentions: analysis of a randomized block Kaczmarz method. Linear Algebra Appl. 2014;441:199–221. https://doi.org/10.1016/j.laa.2012.12.022
9. Dumitrescu B. On the relation between the randomized extended Kaczmarz algorithm and coordinate descent. BIT Numer Math. 2015;55(4):1005–1015. https://doi.org/10.1007/s10543-014-0526-9
10. Needell D, Zhao R, Zouzias A. Randomized block Kaczmarz method with projection for solving least squares. Linear Algebra Appl. 2015;484:322–343. https://doi.org/10.1016/j.laa.2015.06.027
11. Ma A, Needell D, Ramdas A. Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods. SIAM J Matrix Anal Appl. 2015;36(4):1590–1604. https://doi.org/10.1137/15M1014425
12. Gower RM, Richtárik P. Randomized iterative methods for linear systems. SIAM J Matrix Anal Appl. 2015;36(4):1660–1690. https://doi.org/10.1137/15M1025487
13. Oswald P, Zhou W. Convergence analysis for Kaczmarz-type methods in a Hilbert space framework. Linear Algebra Appl. 2015;478:131–161. https://doi.org/10.1016/j.laa.2015.03.028
14. Liu J, Wright SJ. An accelerated randomized Kaczmarz algorithm. Math Comput. 2016;85:153–178. https://doi.org/10.1090/mcom/2971
15. Needell D, Srebro N, Ward R. Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. Math Program. 2016;155(1–2):549–573. https://doi.org/10.1007/s10107-015-0864-7
16. Hefny A, Needell D, Ramdas A. Rows versus columns: randomized Kaczmarz or Gauss-Seidel for ridge regression. SIAM J Sci Comput. 2017;39(5):S528–S542. https://doi.org/10.1137/16M1077891
17. Loizou N, Richtárik P. Momentum and stochastic momentum for stochastic gradient, newton, proximal point and subspace descent methods. 2017 [cited 2017 Dec 22]. Available from: https://arxiv.org/pdf/1712.09677
18. Ma A, Needell D, Ramdas A. Iterative methods for solving factorized linear systems. SIAM J Matrix Anal Appl. 2018;39(1):104–122. https://doi.org/10.1137/17M1115678
19. Bai ZZ, Wu WT. On greedy randomized Kaczmarz method for solving large sparse linear systems. SIAM J Sci Comput. 2018;40(1):A592–A606. https://doi.org/10.1137/17M1137747
20. McCormick SF. The methods of Kaczmarz and row orthogonalization for solving linear equations and least squares problems in Hilbert space. Indiana Univ Math J. 1977;26(6):1137–1150. https://doi.org/10.1512/iumj.1977.26.26090
21. Mandel J. Convergence of the cyclical relaxation method for linear inequalities. Math Program. 1984;30(2):218–228. https://doi.org/10.1007/BF02591886
22. Nutini J, Sepehry B, Virani A, Laradji I, Schmidt M, Koepke H. Convergence rates for greedy Kaczmarz algorithms. Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI); 2016 Jun 25–29; Jersey City, NJ. Corvallis, OR: AUAI Press; 2016.
23. Popa C. Convergence rates for Kaczmarz-type algorithms. Numer Algoritm. 2018;79(1):1–17.
24. Bai ZZ, Wu WT. On relaxed greedy randomized Kaczmarz methods for solving large sparse linear systems. Appl Math Lett. 2018;83:21–26. https://doi.org/10.1016/j.aml.2018.03.008
25. Bai ZZ, Wu WT. On convergence rate of the randomized Kaczmarz method. Linear Algebra Appl. 2018;553:252–269. https://doi.org/10.1016/j.laa.2018.05.009
26. Haddock J, Needell D. On Motzkin's method for inconsistent linear systems. BIT Numer Math. 2018:1–15.
27. Ben-Israel A, Greville TNE. Generalized inverses: Theory and applications. New York, NY: Springer; 2003. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC.
28. Popa C. Least-squares solution of overdetermined inconsistent linear systems using Kaczmarz's relaxation. Int J Comput Math. 1995;55(1–2):79–89.
29. Popa C. Extensions of block-projections methods with relaxation parameters to inconsistent and rank-deficient least-squares problems. BIT Numer Math. 1998;38(1):151–176. https://doi.org/10.1007/BF02510922
30. Popa C. Characterization of the solutions set of inconsistent least-squares problems by an extended Kaczmarz algorithm. Korean J Comput Appl Math. 1999;6(1):51–64.
31. Davis TA, Hu Y. The University of Florida sparse matrix collection. ACM Trans Math Softw. 2011;38(1). Article 1. https://doi.org/10.1145/2049662.2049663
32. Zouzias A, Freris NM. Randomized gossip algorithms for solving Laplacian systems. Paper presented at: 2015 European Control Conference (ECC); 2015 Jul 15–17; Linz, Austria. Piscataway, NJ: IEEE; 2015.
33. Loizou N, Richtárik P. A new perspective on randomized gossip algorithms. Paper presented at: 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2016 Dec 7–9; Washington, DC. Piscataway, NJ: IEEE; 2016.
34. Loizou N, Richtárik P. Accelerated gossip via stochastic heavy ball method. Paper presented at: 56th Annual Allerton Conference on Communication, Control, and Computing; 2018 Oct 2–5; Monticello, IL. Piscataway, NJ: IEEE; 2018.

How to cite this article: Du K. Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss–Seidel algorithms. *Numer Linear Algebra Appl.* 2019;e2233. https://doi.org/10.1002/nla.2233