On randomized Kaczmarz type methods for solving large scale systems of ill-posed equations

J.C. Rabelo† Y. Saporito‡ A. Leitão§

† Department of Mathematics, Federal University of Piaui, 64049-550 Teresina, Brazil.
‡ EMAp, Getulio Vargas Fundation, Praia de Botafogo 190, 22250-900 Rio de Janeiro, Brazil.
§ Dep. of Mathematics, Federal Univ. of St. Catarina, P.O. Box 476, 88040-900 Florianópolis, Brazil.

December 15, 2020

Abstract

In this article we investigate a family of *randomized Kaczmarz type methods*, for solving systems of linear ill-posed equations. The method under consideration is a probabilistic version of the projective Landweber-Kaczmarz (PLWK) method in [20] (see also [21]). We prove mean square convergence to zero of the iteration error.

Numerical tests are presented for a linear ill-posed problem modeled by a linear system with over 10^7 equations, indicating a superior performance of the proposed method when compared with other well established iterations. Our preliminary investigation indicates that the proposed iteration is a promising alternative for computing stable solutions of large scale systems of linear ill-posed equations.

Keywords. Ill-posed problems; Linear systems; Landweber-Kaczmarz method, Randomized method.

AMS Classification: 65J20, 47J06.

1 Introduction

The classical Kaczmarz iteration consisting of cyclic orthogonal projections was devised in 1937 by the Polish mathematician Stefan Kaczmarz for solving systems of linear equations [17]. This method is simple to implement, and it was successfully used for solving ill-posed linear systems related to several relevant applications, e.g., X-ray Tomography[1] [15, 16, 24, 25, 26, 27] and Signal Processing [6, 28, 34].

The starting point of our approach is the projective Landweber (PLW) method [21] and its corresponding Kaczmarz version, the projective Landweber-Kaczmarz (PLWK) method [20]. Our main goal is to modify the PLWK, in order to obtain an efficient method for computing (in a stable way) approximate solutions to large scale systems of linear ill-posed equations.

1.1 Ill-posed systems:

The *inverse problem* we are interested in consists of determining an unknown quantity $x \in X$ from the set of data $(y_0, \ldots, y_{N-1}) \in Y^N$, where X, Y are Hilbert spaces and $N >> 1$ is large

† joelrabelo@ufpi.edu.br, ‡ yuri.saporito@gmail.com, § acgleitao@gmail.com.

[1] In the tomography literature, the Kaczmarz method is called "Algebraic Reconstruction Technique" (ART).

(the case $y_i \in Y_i$ with possibly different spaces $Y_0, \dots, Y_{N-1}$ can be treated analogously). In practical situations, the exact data are not known. Instead, only approximate measured data $y_i^\delta \in Y$ are available such that

$$\|y_i^\delta - y_i\| \ \le\ \delta_i\,, \quad i = 0, \dots, N-1\,, \tag{1}$$

with $\delta_i > 0$ (noise level). We use the notation $\delta := (\delta_0, \dots, \delta_{N-1})$.

The finite set of data above is obtained by indirect measurements of the parameter x, this process being described by the model $A_i x = y_i$, for $i = 0, \dots, N-1$. Here $A_i : X \to Y$ are linear ill-posed operators [10]. Summarizing, the abstract formulation of the inverse problems under consideration consists in finding $x \in X$ such that

$$A_i\, x \ =\ y_i^\delta\,, \quad i = 0, \dots, N-1\,. \tag{2}$$

Standard methods for the solution of system (2) are based in the use of *Iterative type regularization* [1, 9, 14, 18, 19] or *Tikhonov type regularization* [9, 23, 30, 32, 33, 29] after rewriting (2) as a single equation

$$\mathbf{A}x = \mathbf{y}^\delta, \qquad \text{with} \qquad \mathbf{A} := (A_0, \dots, A_{N-1}) : X \to Y^N, \quad \mathbf{y}^\delta := \big(y_0^\delta, \dots, y_{N-1}^\delta\big). \tag{3}$$

If one resorts to the functional analytical formulation (3), one has to face the numerical challenges of solving a large scale system of ill-posed equations [7]. When applied to (3), the above mentioned solution methods may become inefficient if N is large.

An alternative technique for solving system (2) in a stable way is to use *Kaczmarz (cyclic) type regularization methods*. This technique was introduced in [13, 11], [8], [12], [2], [22] and [5] for the Landweber iteration, the Steepest-Descent iteration, the Expectation-Maximization iteration, the Levenberg-Marquardt iteration, the REGINN-Landweber iteration, and the Iteratively Regularized Gauss-Newton iteration respectively.

1.2 PLW and PLWK methods:

The PLW method [21] (originally proposed for nonlinear ill-posed equations) is an iterative type method for solving (2) when $N = 1$, i.e., $A_0 x = y_0$. This method produces a sequence (x_k^δ) such that, in each iteration k, a half space $H_{x_k^\delta} := \{z \in X,\ \langle z - x_k^\delta, A_0^*(y^\delta - A_0 x_k^\delta)\rangle \ge \|y^\delta - A_0 x_k^\delta\|^2\}$ separating the current iterate x_k^δ from the solution set $A_0^{-1}(y_0)$ is defined, and x_{k+1}^δ (the next iterate) is a relaxed projection of x_k^δ onto this set. The resulting iterative method for solving $A_0 x = y_0^\delta$ can be written in the form

$$x_{k+1}^\delta \ :=\ x_k^\delta \ -\ \theta_k\, \lambda_k\, A_0^*\big(A_0 x_k^\delta - y_0^\delta\big)\,, \tag{4}$$

where $\theta_k \in (0, 2)$ is a relaxation parameter and $\lambda_k \ge 0$ gives the exact projection of x_k^δ onto $H_{x_k^\delta}$ (see [21, Eq. (8)]). Observe that this iteration is a Landweber iteration with a stepsize control.

The PLWK method [20] (originally proposed for systems of nonlinear ill-posed equations) is an iterative method for solving (2) when $N > 1$. It consists in coupling the PLW method (4) with the Kaczmarz (cyclic) strategy and incorporating a bang-bang parameter, namely

$$x_{k+1}^\delta \ :=\ x_k^\delta \ -\ \theta_k\, \lambda_k\, \omega_k\, A_{[k]}^*\big(A_{[k]} x_k^\delta - y_{[k]}^\delta\big)\,. \tag{5a}$$

Here the parameters θ_k, λ_k have the same meaning as in (4), while

$$\omega_k \ =\ \omega_k(\delta_{[k]}, y_{[k]}^\delta) \ :=\ \begin{cases} 1 & \|A_{[k]} x_k^\delta - y_{[k]}^\delta\| > \tau \delta_{[k]} \\ 0 & \text{otherwise} \end{cases}\,, \tag{5b}$$

where $\tau > 1$ is an appropriate chosen positive constant and $[k] := (k \mod N) \in \{0, \ldots, N-1\}$.

As usual in Kaczmarz type algorithms, a group of N subsequent steps (starting at some integer multiple of N) is called a cycle. In the case of noisy data, the iteration terminates if all ω_k become zero within a cycle, i.e., if $\|A_i x^\delta_{k+i} - y^\delta_i\| \le \tau\delta_i$, $i \in \{0, \ldots, N-1\}$, for some integer multiple k of N.

In [20] the authors also consider the PLWKr method, namely a randomized version of the PLWK method (in the spirit of [3]) where $[k]$ is randomly chosen in $\{0, \ldots, N-1\}$ (the cyclic structure of PLWK is preserved, i.e., within a cycle each equation is chosen exactly once).

The PLWK iteration in (5) exhibits the following characteristic: For noise free data, $\omega_k = 1$ for all k and each cycle consist of exactly N steps of type (4). Thus, the numerical effort required for the computation of one cycle of PLWK rivals the effort needed to compute one step of PLW (or Landweber) for (3). In this manuscript we propose and analyze a randomized version of the PLWK method, namely the *randomized projective Landweber-Kaczmarz* (rPLWK) method. Our main goal is to modify the PLWK, in order to obtain an efficient method for computing approximate solutions to large scale systems (2) of ill-posed equations.

1.3 Outline of the manuscript:

In Section 2 we state the main assumptions and introduce the rPLWK method.
Section 3 is devoted to the convergence analysis of rPLWK. We estimate the *average gain* (Proposition 3.2), prove monotonicity of *average iteration error* (Corollary 3.4) and square summability of the *average residuals* (Corollary 3.5). Moreover, convergence for exact data is proven (Theorem 3.7).
In Section 4 we present numerical experiments for a linear ill-posed problem modeled by a Hilbert type matrix with over 10^6 lines, while Section 5 is devoted to final remarks and conclusions.

2 The randomized PLWK method:

In what follows we introduce the *randomized projective Landweber-Kaczmarz* (rPLWK) method for solving the linear ill-posed problem (1), (2) in the case of exact data, i.e., $\delta_i = 0$. In this case, the inverse problem can be written in the form

$$A_i\, x \;=\; y_i\,, \quad i = 0, \ldots, N-1\,, \tag{6}$$

or simply $\mathbf{A}\, x = \mathbf{y}$ (compare with (3)).

We start this section by presenting the main assumptions, which are required for the analysis derived in this paper.

2.1 Main assumptions

We assume that some guess $x_0 \in X$ for the solution of (6) is given (e.g., $x_0 = 0$) as well as a sequence sequence $(\theta_k) \in \mathbb{R}$ of relaxation parameters. For the remaining of this article we suppose that the following assumptions hold true:

- (A1) There exists $x^\star \in X$ s.t. $A_i\, x^\star = y_i$, $i = 0, \ldots, N-1$; here $y_i \in R(A_i)$ are exact data.
- (A2) $A_i : X \to Y$ are linear, bounded and ill-posed operators, i.e., even if the operator $A_i^{-1} : R(A_i) \to X$ (the left inverse of A_i) exists, it is not continuous.
- (A3) The sequence (θ_k) satisfies $0 < \inf_k \theta_k$ and $\sup_k \theta_k < 2$.

Notice that Assumption (A2) implies the existence of a constant $C > 0$ s.t. $\max_i \|A_i\| \le C$.

2.2 Description of the method

In the sequel we introduce the rPLWK method for solving (6). Given x_0 and (θ_k) as in Section 2.1, we consider the sequence $(x_k) \in X$ genarated by the iteration formula

$$x_{k+1} \; = \; x_k \; - \; \theta_k \, \lambda_{I_k} \, A^*_{I_k} (A_{I_k} x_k - y_{I_k}) \, , \; k = 0, 1, \dots \tag{7a}$$

where the stepsize λ_{I_k} is given by

$$\lambda_{I_k} \; := \; \begin{cases} \dfrac{\|A_{I_k} x_k - y_{I_k}\|^2}{\|A^*_{I_k}(A_{I_k} x_k - y_{I_k})\|^2} & , \text{ if } A^*_{I_k}(A_{I_k} x_k - y_{I_k}) \neq 0 \\ 0 & , \text{ otherwise} \end{cases} \tag{7b}$$

Here I_k is an independent and identically distributed sequence of indexes taking values in $\{0, \dots, N-1\}$ in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote $p_i = \mathbb{P}(I_k = i)$.

The careful reader observes that, differently from classical Kaczmarz type methods (e.g., Kaczmarz/ART [17], LWK [13, 11], PLW [21], PLWK and its random variant PLKWr [20], LMK [2], EMK [12]), the rPLWK method exhibits no cyclic structure since the choice of the index I_k is independent of I_j for $j = 1, \dots, k-1$. probabilistic The stochastic structure of the rPLKW method is motivated by the ideas discussed in [31], where the operator $\mathbf{A}$ in (3) is considered to be of the form $\mathbf{A} = (A_i)_{i=0}^{N-1} \in \mathbb{R}^{N \times M}$, i.e., a matrix with lines $A_i \in \mathbb{R}^{1,M}$, $X = \mathbb{R}^{M,1}$, $Y = \mathbb{R}$ and $N >> M$. The authors propose a non-cyclic method with an iterative step analog to the step of the Kaczmarz method.[2] In [31] the index I_k is chosen from the set $\{0, \dots, N-1\}$ at random, with probability p_i proportional to $\|A_i\|^2$.

Remark 2.1. *The* **rPLWK method with exact projections** *is obtained by taking $\theta_k = 1$ in (7a), which amounts to define x_{k+1} as the orthogonal projection of x_k onto H_{I_k, x_k}, where*

$$H_{i,x} := \{ z \in X \mid \langle z - x, \; A^*_i (y_i - A_i x) \rangle \geq \|y_i - A_i x\|^2 \}$$

(compare with the PLW method in [21]). A relaxed variant of the rPLWK method uses $\theta_k \in (0, 2)$ so that x_{k+1} is defined as a relaxed projection of x_k onto H_{I_k, x_k}.

Remark 2.2. *Notice that $A^*_i(A_i x - y_i) = 0$ implies $A_i x = y_i$.[3] Consequently, the rPLWK method in (7) can be interpreted as follows:*

- *If $A_{I_k} x_k \neq y_{I_k}$, then x_{k+1} is given by (7a) with $\lambda_k = \|A_{I_k} x_k - y_{I_k}\|^2 \|A^*_{I_k}(A_{I_k} x_k - y_{I_k})\|^{-2}$;*
- *If $A_{I_k} x_k = y_{I_k}$, then $x_{k+1} = x_k$ and $\lambda_k = 0$.*

Remark 2.3. *Assumption (A2) imply $\lambda_{I_k} \geq C^{-2}$ if $\|A_{I_k} x_k - y_{I_k}\| > 0$ (see also Remark 2.2). In other words, C^{-2} is a natural lower bound for the stepsizes defined in (7b), whenever x_k is not a solution of the equation $A_{I_k} x = y_{I_k}$.*

3 Convergence Analysis

In what follows we estimate the "average gain" $\mathbb{E}[\|x^* - x_{k+1}\|^2] - \mathbb{E}[\|x^* - x_k\|^2]$, where $x^* \in X$ is a solution of (6). This is a fundamental result for the forthcoming analysis.

Remark 3.1. *Given $x^* \in X$ a solution of (6), the mean square iteration error $\mathbb{E}[\|x^* - x_k\|^2]$ is defined by the average error over all possible realizations of $I_1, \dots, I_{k-1}$ that define x_k.*

[2] Namely, $x_{k+1} = x_k - (y_{I_k} - A_{I_k} x_k) \|A_{I_k}\|^{-2} A^*_{I_k}$, $k = 0, 1, \dots$

[3] Indeed, notice that $A_i x - y_i \in R(A_i)$. Moreover, $A^*_i(A_i x - y_i) = 0$ implies $A_i x - y_i \in N(A^*_i) = R(A_i)^\perp$. Consequently, $A_i x - y_i \in R(A_i) \cap R(A_i)^\perp = \{0\}$.

Proposition 3.2. *Let assumptions (A1), (A2) hold true and* (x_k) *be a sequence generated by the rPLWK method* (7). *Then, for any* x^* *solution of* (6) *we have*

$$\mathbb{E}[\|x^* - x_{k+1}\|^2] - \mathbb{E}[\|x^* - x_k\|^2] \;=\; \theta_k(\theta_k - 2)\,\mathbb{E}[\lambda_{I_k}\|A_{I_k}x_k - y_{I_k}\|^2]\,,\;\; k = 0, 1, \ldots \tag{8}$$

Proof. From (A1) we know that $\bigcap_i A_i^{-1}(y_i) \neq \emptyset$. Thus, for x^* as in the assumptions we have

$$\begin{aligned}
\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2 &= 2\langle x^* - x_k, x_k - x_{k+1}\rangle + \|x_k - x_{k+1}\|^2 \\
&= -2\theta_k\lambda_{I_k}\langle x_k - x^*, A_{I_k}^*(A_{I_k}x_k - y_{I_k})\rangle + \theta_k^2\lambda_{I_k}^2\|A_{I_k}^*(A_{I_k}x_k - y_{I_k})\|^2 \\
&= -2\theta_k\frac{\|A_{I_k}x_k - y_{I_k}\|^4}{\|A_{I_k}^*(A_{I_k}x_k - y_{I_k})\|^2} + \theta_k^2\frac{\|A_{I_k}x_k - y_{I_k}\|^4}{\|A_{I_k}^*(A_{I_k}x_k - y_{I_k})\|^2} \\
&= \theta_k(\theta_k - 2)\frac{\|(A_{I_k}x_k - y_{I_k})\|^4}{\|A_{I_k}^*(A_{I_k}x_k - y_{I_k})\|^2} \;=\; \theta_k(\theta_k - 2)\lambda_{I_k}\|A_{I_k}x_k - y_{I_k}\|^2.
\end{aligned} \tag{9}$$

Thus, if we denote by $\mathcal{F}_k$ the σ-algebra generated by $I_1, \ldots, I_{k-1}$, we conclude that x_k is measurable with respect to $\mathcal{F}_k$, and I_k is independent of it. Consequently,

$$\mathbb{E}[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2|\mathcal{F}_k] \;=\; \theta_k(\theta_k - 2)\mathbb{E}[\lambda_{I_k}\|A_{I_k}x_k - y_{I_k}\|^2|\mathcal{F}_k].$$

Now, taking the full conditional yields (8). ☐

Remark 3.3. *Notice that, in the final step of the proof of Proposition 3.2 we have*

$$\mathbb{E}[\lambda_{I_k}\|A_{I_k}x_k - y_{I_k}\|^2|\mathcal{F}_k] \;=\; \sum_{i=0}^{N-1} p_i\,\lambda_i\,\|A_ix_k - y_i\|^2 \;=\; \sum_{i=0}^{N-1} p_i\,\frac{\|A_ix_k - y_i\|^4}{\|A_i^*(A_ix_k - y_i)\|^2}\,.$$

A direct consequence of Proposition 3.2 is the monotonicity of the mean square iteration error:

Corollary 3.4. *Let assumption (A1) hold true and* (x_k) *be a sequence generated by the rPLWK method* (7). *Then, for any* x^* *solution of* (6) *we have*

$$\mathbb{E}[\|x^* - x_{k+1}\|^2] \;\leq\; \mathbb{E}[\|x^* - x_k\|^2]\,,\;\; k = 0, 1, \ldots \tag{10}$$

Another consequence of Proposition 3.2 is discussed in the next corollary. This result is needed for the proof of the main convergence theorem of this manuscript (see Theorem 3.7).

Corollary 3.5. *Let assumptions (A1), (A3) hold true and* (x_k) *be a sequence generated by the rPLWK method* (7). *Then, the series*

$$\sum_{k=0}^{\infty} \theta_k\,\mathbb{E}[\lambda_{I_k}\|A_{I_k}x_k - y_{I_k}\|^2] \quad \textit{and} \quad \sum_{k=0}^{\infty} \mathbb{E}[\|A_{I_k}x_k - y_{I_k}\|^2]$$

are summable.

Proof. The summability of the first series follows from (A3) together with the fact that the series $\sum_k \theta_k(2 - \theta_k)\mathbb{E}[\lambda_{I_k}\|A_{I_k}x_k - y_{I_k}\|^2]$ is summable (see Proposition 3.2). The summability of the second series follows from (A3), the summability of the first series, and the facts:

i) $\lambda_{I_k} = 0$ iff $A_{I_k}x_k = y_{I_k}$; *ii)* $\lambda_{I_k} \geq 1/C^2$ whenever $\|A_{I_k}x_k - y_{I_k}\| > 0$
(see Remarks 2.2 and 2.3). ☐

Yet another consequence of Proposition 3.2 is the fact that the sequence (x_k) generated by the rPLKW method with exact projections (i.e., obtained by taking $\theta_k = 1$ in (7a)) is an **average reasonable wanderer** in the sense of [4], i.e., $\sum_{k=0}^{\infty} \mathbb{E}[\|x_k - x_{k+1}\|^2] < \infty$. Indeed, since $\theta_k = 1$, it follows from (7) that either $A_{I_k} x_k = y_{I_k}$ and $x_{k+1} = x_k$; or $\|A_{I_k} x_k - y_{I_k}\| > 0$ and $x_{k+1} = x_k - \lambda_{I_k} A^*_{I_k}(A_{I_k} x_k - y_{I_k}) \in H_{I_k, x_k}$ (see Remark 2.1). In either case we have $\langle x_k - x^*, x_k - x_{k+1}\rangle = \|x_k - x_{k+1}\|^2$ for any solution x^* of $A_{I_k} x = y_{I_k}$.[4] Thus, arguing as in (9) we obtain

$$\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2 \;=\; 2\langle x^* - x_k, x_k - x_{k+1}\rangle + \|x_k - x_{k+1}\|^2 \;=\; -\|x_k - x_{k+1}\|^2 .$$

Now, arguing as in Corollary 3.5 we conclude that $\sum\limits_{k=0}^{\infty} \mathbb{E}[\|x_k - x_{k+1}\|^2] < \infty$.

We are now ready to state and prove the main result of this manuscript, namely convergence in mean square of the rPLWK method. First, however, we briefly recall the concept of **minimal norm solutions** of (6).

Remark 3.6. *It is worth noticing that there exists an x_0-minimal norm solution of (6), i.e., a solution $x^\dagger$ of (6) satisfying $\|x^\dagger - x_0\| = \inf\{\|x - x_0\|;\ x \in X$ is solution of (6)$\}$.*[5] *Moreover, $x^\dagger$ is the only solution of (6) with this property.*

Theorem 3.7 (Convergence for exact data)**.** *Let assumptions (A1), ...,(A3) hold true. Then, any sequence (x_k) generated by the rPLWK method (7) converges in mean square to $x^\dagger$, the x_0-minimal norm solution of (6). I.e., $\mathbb{E}[\|x^\dagger - x_k\|^2] \to 0$ as $k \to \infty$.*

Proof. Let $x^\star$ be given as in (A1). The proof is divided in three main steps:
Step 1. We prove that (x_k) is a Cauchy sequence.
It is enough to prove that $e_k := x^\star - x_k$ is a Cauchy sequence. From Corollary 3.4 follows

$$\lim_{k\to\infty} \mathbb{E}[\|e_k\|^2] \;=\; \varepsilon\,, \tag{11}$$

for some $\varepsilon \geq 0$. In order to prove that (e_k) is a Cauchy sequence, we first prove

$$\mathbb{E}[\langle e_n - e_k, e_n\rangle] \to 0 \quad \text{and} \quad \mathbb{E}[\langle e_l - e_n, e_n\rangle] \to 0 \quad \text{as} \quad k,\ l \to \infty, \tag{12}$$

with $k \leq l$ for some $k \leq n \leq l$ (compare with [14, Theorem 2.3]).Notice that $\mathbb{E}[\langle \cdot\,, \cdot\rangle]$ defines an inner product in $L^2(\Omega; X)$.[6]

Notice that, for any $k \leq l$, one can always choose an index n with $k \leq n \leq l$ such that

$$\mathbb{E}[\lambda_I \|A_I\, x_n - y_I\|^2] \;\leq\; \mathbb{E}[\lambda_I \|A_I\, x_j - y_I\|^2]\,, \quad \forall\ k \leq j \leq l \tag{13}$$

holds true.[7] Next, we argue with (7a) and the Cauchy–Schwartz inequality to estimate

$$\begin{aligned}
\big|\mathbb{E}[\langle e_n - e_k, e_n\rangle]\big| \;&=\; \Big|\sum_{j=k}^{n-1} \mathbb{E}[\langle x_{j+1} - x_j, x^\star - x_n\rangle]\Big| \\
&=\; \Big|\sum_{j=k}^{n-1} \mathbb{E}[\theta_j\, \lambda_{I_j} \langle A^*_{I_j}(y_{I_j} - A_{I_j} x_j), x^\star - x_n\rangle]\Big| \\
&=\; \Big|\sum_{j=k}^{n-1} \theta_j\, \mathbb{E}[\lambda_I \langle y_I - A_I x_j, A_I(x^\star - x_n)\rangle]\Big| \\
&\leq\; \sum_{j=k}^{n-1} \theta_j\, \mathbb{E}[\lambda_I \|A_I x_j - y_I\|^2]^{\frac{1}{2}}\ \mathbb{E}[\|A_I x_n - y_I\|^2]^{\frac{1}{2}} .
\end{aligned} \tag{14}$$

[4] Notice that all solutions of $A_{I_k} x = y_{I_k}$ belong to H_{I_k, x_k}.
[5] See, e.g., [9] for details.
[6] $L^2(\Omega; X)$ is the space of square integrable random variables defined on Ω and taking values in X.
[7] We adopt the notation $\mathbb{E}[\lambda_I \|A_I\, x_k - y_I\|^2] = \mathbb{E}[\lambda_{I_k} \|A_{I_k}\, x_k - y_{I_k}\|^2]$.

Now notice that, due to (13), we have for all $j \in \{k, k+1, \ldots, l\}$

$$\mathbb{E}[\|A_I x_n - y_I\|^2] = C^2\mathbb{E}[C^{-2}\|A_I x_n - y_I\|^2] \le C^2\mathbb{E}[\lambda_I\|A_I x_n - y_I\|^2] \le C^2\mathbb{E}[\lambda_I\|A_I x_j - y_I\|^2].$$

Substituting this inequality in (14) we obtain

$$\left|\mathbb{E}[\langle e_n - e_k, e_n\rangle]\right| \le C\sum_{j=k}^{n-1}\theta_j\,\mathbb{E}[\lambda_I\|A_I x_j - y_I\|^2] = C\sum_{j=k}^{n-1}\theta_j\,\mathbb{E}[\lambda_{I_j}\|A_{I_j}x_j - y_{I_j}\|^2].$$

Consequently, Corollary 3.5 allow us to conclude $\mathbb{E}[\langle e_n - e_k, e_n\rangle] \to 0$ as $k, l \to \infty$. Analogously one proves $\mathbb{E}[\langle e_l - e_n, e_n\rangle] \to 0$ as $k, l \to \infty$, establishing (12).

Finally, one argues with (12), (11), inequality $\mathbb{E}[\|e_j - e_k\|^2]^{\frac12} \le \mathbb{E}[\|e_j - e_l\|^2]^{\frac12} + \mathbb{E}[\|e_l - e_k\|^2]^{\frac12}$ and identities

$$\begin{aligned}\mathbb{E}[\|e_j - e_l\|^2] &= 2\mathbb{E}[\langle e_l - e_j, e_l\rangle] + \mathbb{E}[\|e_j\|^2] - \mathbb{E}[\|e_l\|^2],\\ \mathbb{E}[\|e_l - e_k\|^2] &= 2\mathbb{E}[\langle e_l - e_k, e_l\rangle] + \mathbb{E}[\|e_k\|^2] - \mathbb{E}[\|e_l\|^2]\end{aligned}$$

to conclude that $\mathbb{E}[\|e_j - e_k\|^2] \to 0$, as $k, l \to \infty$, i.e., (e_k) is a Cauchy sequence in $L^2(\Omega; X)$.

Step 2. We prove that (x_k) converges to some x^* in $L^2(\Omega; X)$, which is a solution of (6). Since (x_k) is Cauchy in $L^2(\Omega; X)$, it has an accumulation point x^*. Moreover, it follows from Corollary 3.5 that the mean square residuals $\mathbb{E}[\|A_I x_k - y_I\|^2]$ converge to zero as $k \to \infty$. Consequently, $\mathbb{E}[\|A_I x^* - y_I\|^2] = 0$, i.e., $x^* \in X$ and $\|A_i x^* - y_i\|^2 = 0$ for $i = 0, \ldots, N-1$. Thus x^* is a solution of (6).

Step 3. We prove that $x^* = x^\dagger$.
Indeed, notice that $x_{k+1} - x_k \in \mathcal{R}(A_{I_k}^*) \subset \mathcal{N}(A_{I_k})^\perp \subset \mathcal{N}(\mathbf{A})^\perp$, for $k = 0, 1, \ldots$[8] Thus, an inductive argument shows that $x^* \in x_0 + \mathcal{N}(\mathbf{A})^\perp$. However, $x^\dagger$ is the only solution of (6) with this property (see Remark 3.6), concluding the proof. □

4 Numerical experiments

In this section the rPLWK method in (7) is implemented for solving a benchmark problem, which happens to be a well known system of linear ill-posed equations.[9]

Let $\mathbf{B} = (B_i)_{i=0}^{N-1} \in \mathbb{R}^{N\times M}$ be a Hilbert type matrix with lines $B_i = \left(\frac{1}{i+j+1}\right)_{j=0}^{M-1} \in \mathbb{R}^{1,M}$, $X = \mathbb{R}^{M,1}$ and $Y = \mathbb{R}$, where $N = 10^6$ and $M = 10^2$.

The operator $\mathbf{A} = (A_i)_{i=0}^{N-1} \in \mathbb{R}^{N\times M}$ with lines $A_i \in \mathbb{R}^{1,M}$, is obtained by a random shuffle of the lines of $\mathbf{B}$. In our numerical experiments we set $x^\star = (1, \ldots, 1) \in X$ and compute the corresponding exact data $y_i = A_i x^\star$. The noise levels are $\delta_i = 10^{-16}$, what corresponds to the MATLAB double precision accuracy. The performance of the rPLWK method is compared against two concurrent Kaczmarz type methods, namely: (1) Landweber Kaczmarz with random ordering of equations within cycles [21]; (2) Projective Landweber Kaczmarz with random ordering of equations within cycles [20].

In order to better investigate the behavior of iteration (7), four different runs of the rPLWK method are computed for the same set of data. In the first three runs (run 1, 2 and 3), the indexes I_k are chosen from the set $\{0, \ldots, N-1\}$ at random, with equal probability, i.e., $p_i = N^{-1}$ for $i = 1, \ldots, N-1$. In the last run (run *) each index I_k is chosen from the set $\{0, \ldots, N-1\}$ at random, with probability p_i proportional to $\|A_i\|^2$ (as proposed in [31] for the randomized Kaczmarz iteration). The results are shown in Figures 1 to 4.

[8] Here $\mathbf{A} = (A_i)_{i=0}^{N-1} : X \to Y^N$.

[9] Computations are performed using MATLAB® R2017a, running in a Intel® Core™ i7-3520M CPU .

We consider in our numerical experiments systems of four distinct dimensions, namely: $N = 10^4$ in Figure 1, $N = 10^5$ in Figure 2, $N = 10^6$ in Figure 3 and $N = 10^7$ in Figure 4. The first three runs of the rPLKW method (run 1, 2 and 3) produced similar results for all tested dimensions. This last run of the rPLKW method (run *) delivered the best numerical performance in all tested dimensions. Moreover, the improvement of the numerical performance increases with N.

In each figure, the first plot (TOP) shows the evolution of the **residual** $\|\mathbf{A}x_k - \mathbf{y}\|$, while the second plot (BOTTOM) shows the **relative iteration error** $\|x^\star - x_k\|/\|x^\star\|$. All methods are stopped after $20N$ iterative steps (this corresponds to 20 cycles of methods (1) and (2)). In these plots the x-axis shows the accumulated number of operations, measured by the number of computed 'matrix $\times$ vector' products (see Remark 4.1 below).

It is worth noticing that, in Kaczmarz type methods we have monotonicity of the mean square iteration error $\mathbb{E}[\|x^\star - x_k\|^2]$, see Corollary 3.4. However monotonicity of mean square residual $\mathbb{E}[\|\mathbf{A}x_k - \mathbf{y}\|^2]$ cannot be guaranteed. These two facts can be observed in all Figures.

Remark 4.1. *In Kaczmarz type methods, the computation of an iterative step is avoided (i.e., $x_{k+1} = x_k$) whenever the residual satisfies $\|A_i x_k - y_i\| \le \tau\delta_i$ for some $\tau > 1$. Consequently, the numerical burden of computing a cycle differs from method to method (as well as from cycle to cycle of the same method). Therefore, plotting iteration errors (or residuals) after each cycle does not give a proper comparison of the efficiency of these methods. In our figures the evolution of iteration error and residual are plotted as functions of accumulated number of operations (i.e., computed 'matrix $\times$ vector' products). This allows a fair comparison between these methods, since the number of computed iterates is proportional to the total computational burden of a Kaczmarz type method.*

5 Conclusions

We investigate randomized Landweber-Kaczmarz type methods for computing stable approximate solutions to large scale systems of linear ill-posed operator equations. The main contribution of this article is to propose and analyze, in the case of exact data, a stochastic version of the PLKW method in [20] (see also [21]).

We prove monotonicity of the proposed rPLWK method (Corollary 3.4). Moreover, we provide estimates to the "average gain" $\mathbb{E}[\|x^\star - x_k\|^2] - \mathbb{E}[\|x^\star - x_{k+1}\|^2]$ (Proposition 3.2), as well as a lower bound to the proposed stepsizes λ_{I_k} (Remark 2.3). A convergence proof in the case of exact data is given (Theorem 3.7).

An algorithmic implementation of the rPLWK method is proposed. The resulting rPLWK algorithm is tested for a well known benchmark problem modeled by a large scale Hilbert type matrix, and compared with two well known Kaczmarz type methods, namely LWK and PLWK. The obtained results validate the efficiency of our method.

References

[1] A.B. Bakushinsky and M.Y. Kokurin. *Iterative Methods for Approximate Solution of Inverse Problems*, volume 577 of *Mathematics and Its Applications*. Springer, Dordrecht, 2004.

[2] J. Baumeister, B. Kaltenbacher, and A. Leitão. On levenberg-marquardt-kaczmarz iterative methods for solving systems of nonlinear ill-posed equations. *Inverse Probl. Imaging*, 4(3):335–350, 2010.

[3] H.H. Bauschke and J.M. Borwein. Legendre functions and the method of random bregman projections. *Journal of Convex Analysis*, 4(1):27–67, 1997.

[4] F. Browder and W. Petryshyn. Construction of fixed points of nonlinear mappings in hilbert space. *Journal Math. Anal. Appl.*, 20:197–228, 1967.

[5] M. Burger and B. Kaltenbacher. Regularizing Newton-Kaczmarz methods for nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 44:153–182, 2006.

[6] C.L. Byrne. *Signal processing*. Monographs and Research Notes in Mathematics. CRC Press, Boca Raton, FL, second edition, 2015. A mathematical approach.

[7] M. Cullen, M.A. Freitag, S. Kindermann, and R. Scheichl, editors. *Large Scale Inverse Problems*, volume 13 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, Berlin, 2013. Computational methods and applications in the earth sciences.

[8] A. De Cezaro, M. Haltmeier, A. Leitão, and O. Scherzer. On steepest-descent-Kaczmarz methods for regularizing systems of nonlinear ill-posed equations. *Appl. Math. Comput.*, 202(2):596–607, 2008.

[9] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 1996.

[10] C. W. Groetsch. *Stable Approximate Evaluation of Unbounded Operators*, volume 1894 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2007.

[11] M. Haltmeier, R. Kowar, A. Leitão, and O. Scherzer. Kaczmarz methods for regularizing nonlinear ill-posed equations. II. Applications. *Inverse Probl. Imaging*, 1(3):507–523, 2007.

[12] M. Haltmeier, A. Leitão, and E. Resmerita. On regularization methods of EM-Kaczmarz type. *Inverse Problems*, 25:075008, 2009.

[13] M. Haltmeier, A. Leitão, and O. Scherzer. Kaczmarz methods for regularizing nonlinear ill-posed equations. I. convergence analysis. *Inverse Probl. Imaging*, 1(2):289–298, 2007.

[14] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72:21–37, 1995.

[15] G. T. Herman. A relaxation method for reconstructing objects from noisy X-rays. *Math. Programming*, 8:1–19, 1975.

[16] Gabor T. Herman. *Image reconstruction from projections*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. The fundamentals of computerized tomography, Computer Science and Applied Mathematics.

[17] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull. International de l'Academie Polonaise des Sciences. Lett A*, pages 355–357, 1937.

[18] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

[19] L. Landweber. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.

[20] A. Leitão and B.F. Svaiter. On projective landweber-kaczmarz methods for solving systems of nonlinear ill-posed equations. *Inverse Problems*, 32(1):025004, 2016.

[21] A. Leitão and B.F. Svaiter. On a family of gradient type projection methods for nonlinear ill-posed problems. *Numerical Functional Analysis and Optimization*, 39(2-3):1153–1180, 2018.

[22] F. Margotti, A. Rieder, and A. Leitão. A Kaczmarz version of the `reginn`-Landweber iteration for ill-posed problems in Banach spaces. *SIAM J. Numer. Anal.*, 52(3):1439–1465, 2014.

[23] V.A. Morozov. *Regularization Methods for Ill–Posed Problems.* CRC Press, Boca Raton, 1993.

[24] F. Natterer. Regularisierung schlecht gestellter Probleme durch Projektionsverfahren. *Numerische Mathematik*, 28(3):329–341, 1977.

[25] F. Natterer. *The Mathematics of Computerized Tomography.* B.G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester, 1986.

[26] F. Natterer. Algorithms in tomography. In *State of the art in numerical analysis*, volume 63, pages 503–524, 1997.

[27] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction.* SIAM, Philadelphia, 2001.

[28] H.A. Sabbagh, R.K. Murphy, E.H. Sabbagh, J.C. Aldrin, and J.S. Knopp. *Computational electromagnetics and model-based inversion.* Scientific Computation. Springer, New York, 2013. A modern paradigm for eddy-current nondestructive evaluation.

[29] O. Scherzer. Convergence rates of iterated Tikhonov regularized solutions of nonlinear ill-posed problems. *Numerische Mathematik*, 66(2):259–279, 1993.

[30] T.I. Seidman and C.R. Vogel. Well posedness and convergence of some regularisation methods for non–linear ill posed problems. *Inverse Probl.*, 5:227–238, 1989.

[31] Thomas Strohmer and Roman Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009.

[32] A.N. Tikhonov. Regularization of incorrectly posed problems. *Soviet Math. Dokl.*, 4:1624–1627, 1963.

[33] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-Posed Problems.* John Wiley & Sons, Washington, D.C., 1977. Translation editor: Fritz John.

[34] M. Vetterli, J. Kovacevic, and V.K. Goyal. *Foundations of Signal Processing.* Cambridge University Press, Cambridge, 2014. A modern paradigm for eddy-current nondestructive evaluation.

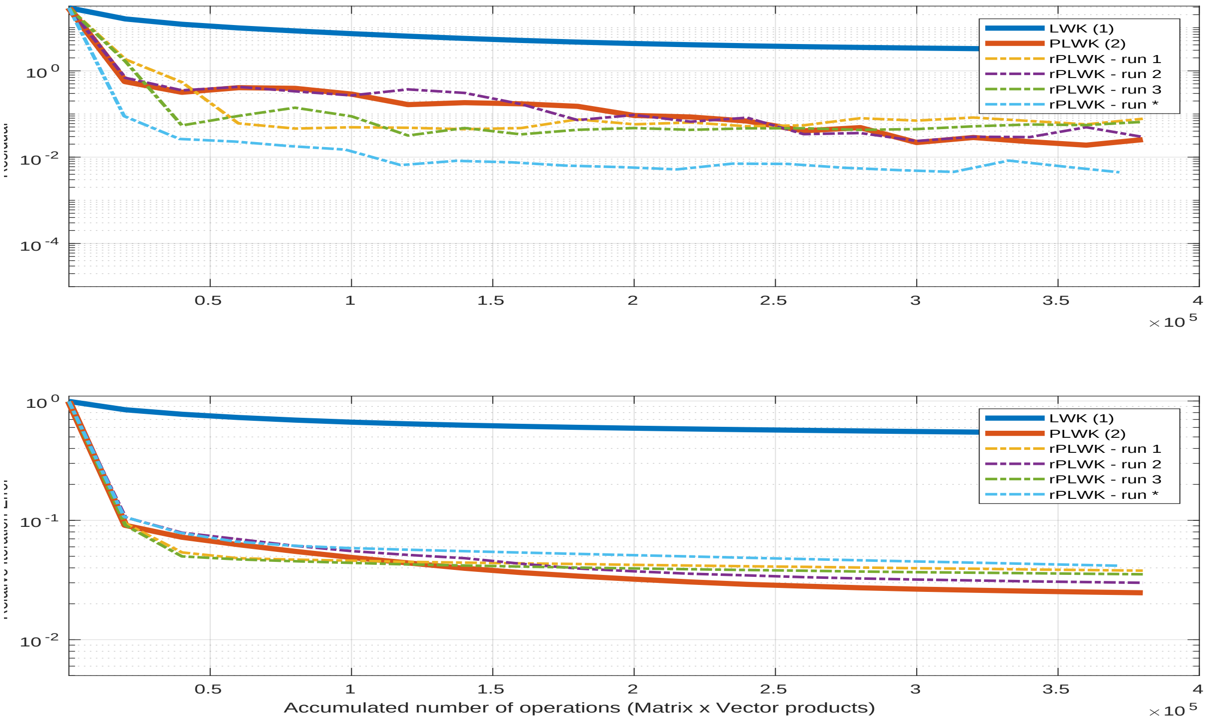

Figure 1: $N = 10^4$: (TOP) Residual $\|\mathbf{A}x_k - y\|$; (BOTTOM) Relative iteration error $\|x^\star - x_k\|/\|x^\star\|$.

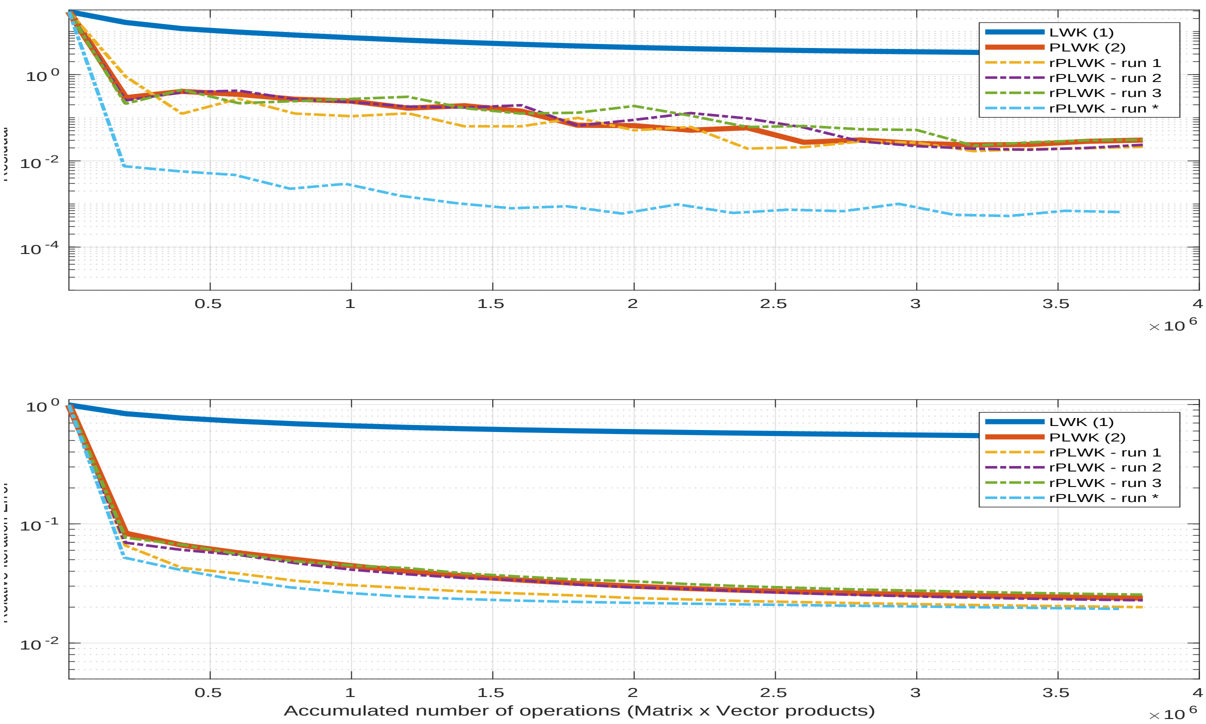

Figure 2: $N = 10^5$: (TOP) Residual $\|\mathbf{A}x_k - y\|$; (BOTTOM) Relative iteration error $\|x^\star - x_k\|/\|x^\star\|$.

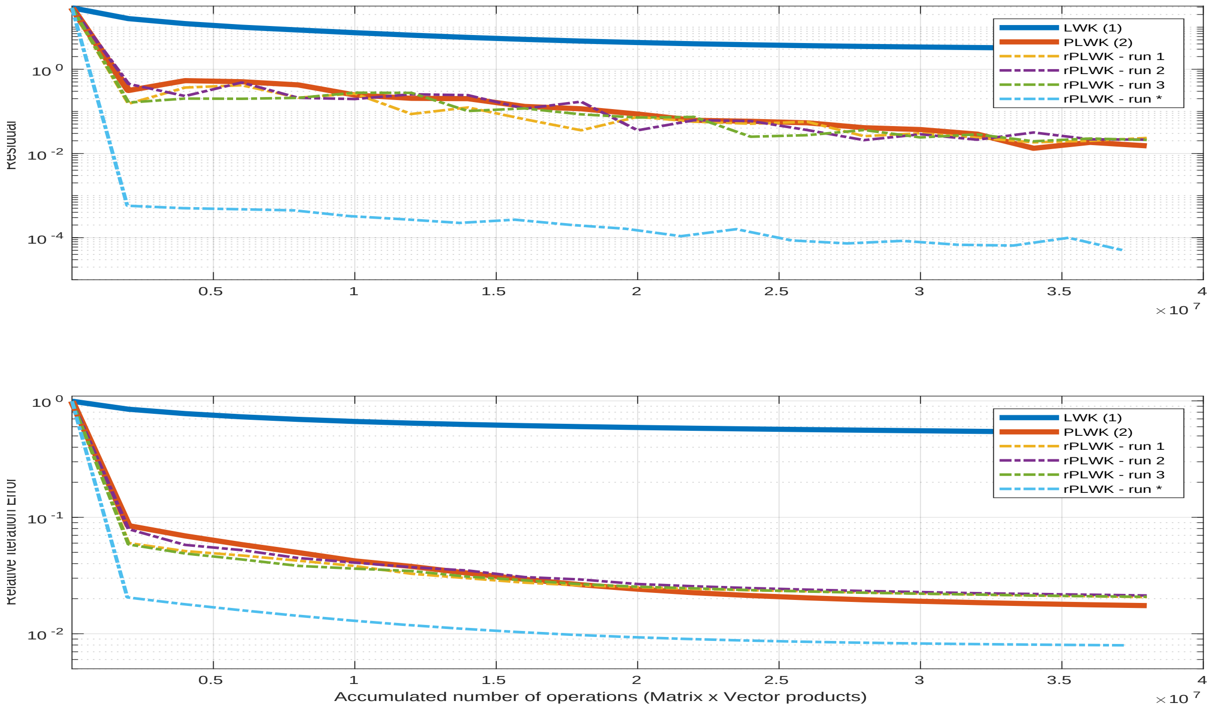

Figure 3: $N = 10^6$: (TOP) Residual $\|\mathbf{A}x_k - y\|$; (BOTTOM) Relative iteration error $\|x^\star - x_k\|/\|x^\star\|$.

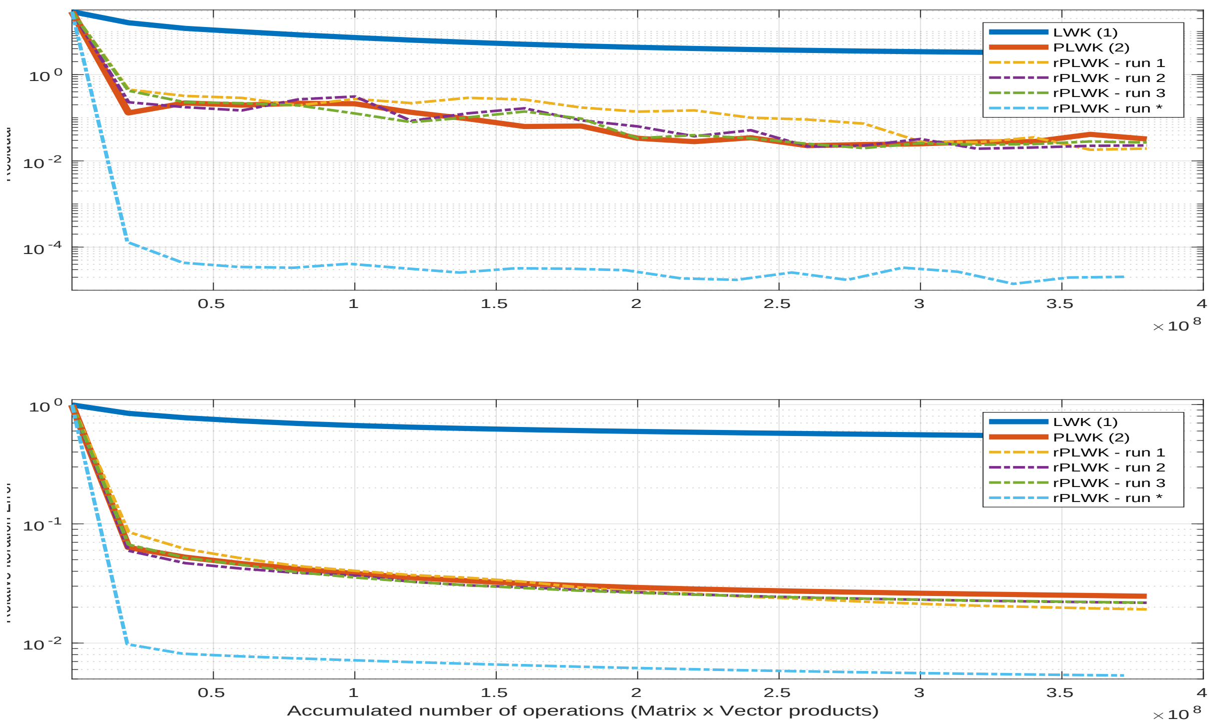

Figure 4: $N = 10^7$: (TOP) Residual $\|\mathbf{A}x_k - y\|$; (BOTTOM) Relative iteration error $\|x^\star - x_k\|/\|x^\star\|$.