Inverse Problems **38** (2022) 025003 (23pp) https://doi.org/10.1088/1361-6420/ac3f80

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

J C Rabelo[1], Y F Saporito[2] and A Leitão[3,*]

[1] Department of Mathematics, Federal University of Piaui, 64049-550 Teresina, Brazil
[2] School of Applied Mathematics, Getulio Vargas Fundation, Praia de Botafogo 190, 22250-900 Rio de Janeiro, Brazil
[3] Department of Mathematics, Federal Univ. of St. Catarina, PO Box 476, 88040-900 Florianópolis, Brazil

E-mail: joelrabelo@ufpi.edu.br, yuri.saporito@fgv.br and acgleitao@gmail.com

Received 14 May 2021, revised 25 November 2021
Accepted for publication 2 December 2021
Published 23 December 2021

Abstract

In this article we investigate a family of *stochastic gradient type methods* for solving systems of linear ill-posed equations. The method under consideration is a stochastic version of the projective Landweber–Kaczmarz method in Leitão and Svaiter (2016 *Inverse Problems* **32** 025004) (see also Leitão and Svaiter (2018 *Numer. Funct. Anal. Optim.* **39** 1153–80)). In the case of exact data, mean square convergence to zero of the iteration error is proven. In the noisy data case, we couple our method with an *a priori* stopping rule and characterize it as a regularization method for solving systems of linear ill-posed operator equations. Numerical tests are presented for two linear ill-posed problems: (i) a Hilbert matrix type system with over 10^8 equations; (ii) a big data linear regression problem with real data. The obtained results indicate superior performance of the proposed method when compared with other well-established random iterations. Our preliminary investigation indicates that the proposed iteration is a promising alternative for computing stable approximate solutions of large scale systems of linear ill-posed equations.

Keywords: ill-posed problems, linear systems, Landweber–Kaczmarz method, stochastic method

(Some figures may appear in colour only in the online journal)

[*] Author to whom any correspondence should be addressed.

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 [20]. This method is simple to implement, and the iterative step can be evaluated with low computational coast. It was successfully used for solving ill-posed linear systems related to several relevant applications, e.g., x-ray tomography [18, 19, 27–30] and signal processing [6, 31, 37].

The starting point of our approach are two Kaczmarz type methods, namely the projective Landweber (PLW) method [24] and its corresponding Kaczmarz version, the projective Landweber–Kaczmarz (PLWK) method [23]. Our main goal is to analyze a stochastic version of the PLWK, aiming to obtain a numerically efficient method for computing stable approximate solutions to large scale systems of linear ill-posed equations.

It is worth noticing that the here proposed stochastic PLWK method can be interpreted as a stochastic gradient descent type method [39–43] for the least-squares problem with a particular choice of the stepsize.

1.1. Problems under consideration

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 $N \gg 1$ is large[4]. In practical situations, the exact data are not known. Instead, only approximate measured data $y_i^\delta \in Y$ are available such that

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

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

The finite set of data above is obtained by indirect measurements of the parameter x, this process being described by the model $y_i = A_i x$, for $i = 0, \ldots, N-1$. Here $A_i : X \to Y$ are linear ill-posed operators [13]. Summarizing, the abstract formulation of the inverse problems under consideration reads: given the data y_i^δ and the levels of noise δ_i as in (1), find an approximate solution to the large scale linear system

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

Standard methods for the solution of (1) and (2) are based in the use of *iterative type regularization* [1, 9, 17, 21, 22] or *Tikhonov type regularization* [9, 26, 32, 33, 35, 36] after rewriting (2) as a single equation

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

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 [2, 5, 8, 14–16, 25] 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.

[4] The case $y_i \in Y_i$ with possibly different spaces $Y_0, \ldots, Y_{N-1}$ can be treated analogously.

1.2. Starting point of our approach: PLW and PLWK methods

The PLW method was originally proposed in [24] for solving nonlinear operator equations (it can be applied for solving (1) and (2) with $N = 1$, i.e., $A_0 x = y_0$ and $\|y_0 - y_0^\delta\| \leqslant \delta$). A sequence (x_k^δ) is generated as follows: at each iteration k, a half space

$$H_{x_k^\delta} := \left\{ z \in X | \langle z - x_k^\delta, A_0^*(y_0^\delta - A_0 x_k^\delta) \rangle \geqslant \|y_0^\delta - A_0 x_k^\delta\| \left(\|y_0^\delta - A_0 x_k^\delta\| - \delta \right) \right\}$$

separating the current iterate x_k^δ from the solution set $A_0^{-1}(y_0)$ is defined; [5] thus, the next iterate x_{k+1}^δ is defined as a (relaxed) orthogonal projection of x_k^δ onto this set. This iterative method can be summarized as follows

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

where $\theta_k \in (0, 2)$ is a relaxation parameter and $\lambda_k \geqslant 0$ gives the exact orthogonal projection of x_k^δ onto $H_{x_k^\delta}$ (see [24, equation (8)] for details). This method corresponds to a Landweber iteration with stepsize defined by (relaxed) orthogonal projections onto the separating sets $H_{x_k^\delta}$.

The PLWK method was originally proposed in [23] for solving systems of nonlinear ill-posed equations as in (1) and (2) when $N > 1$. It consists in coupling the PLW method (4) with the Kaczmarz (cyclic) strategy and incorporating a bang-bang parameter, i.e.,

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

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 $[k] := (k \bmod N) \in \{0, \dots, N-1\}$.

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 $\|A_i x_{k+i}^\delta - y_i^\delta\| \leqslant \tau \delta_i$, $i \in \{0, \dots, N-1\}$, for some integer multiple k of N.

In [23] 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 $\{0, \dots, N-1\}$ (the cyclic structure of PLWK is preserved, i.e., within a cycle each equation is chosen exactly once).

The PLWK iteration (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). However, in the noisy data case, the computational effort for computing a cycle is reduced due to the introduction of the bang-bang parameter ω_k; indeed, the iterative step is not computed if the residual w.r.t. the $[k]$th equation is smaller than $\tau \delta_{[k]}$.

1.3. Main goals

In this manuscript we propose and analyze a stochastic version of the PLWK method, namely the *stochastic projective Landweber–Kaczmarz* (sPLWK) method. Our main goal is to modify

[5] By saying that $H_{x_k^\delta}$ **separates** x_k^δ from $A_0^{-1}(y_0)$ we mean that $A_0^{-1}(y_0) \subset H_{x_k^\delta}$, while $x_k^\delta \notin H_{x_k^\delta}$.

the PLWK, in order to obtain an efficient method for computing stable approximate solutions to large scale systems of ill-posed operator equations (1) and (2). Differently from [23] we propose here a stochastic (noncyclic) method based on the iteration (5), which uses an *a priori* stopping rule in the case of noisy data.

1.4. Outline of the manuscript

In section 2 we state the main assumptions and introduce the sPLWK method. Section 3 is devoted to the analysis of sPLWK in the exact data case. We estimate the *average gain* (lemma 3.2), prove monotonicity of *average iteration error* (proposition 3.3) as well as square summability of the *average residuals* (proposition 3.4). Moreover, convergence for exact data is proven (theorem 3.6). Section 4 is dedicated to regularization properties of sPLWK. The main results are a stochastic stability result (theorem 4.3) and a semi-convergence result (theorem 4.4). In section 5 we present numerical experiments. Two distinct applications are considered: in section 5.1 a linear ill-posed problem modeled by a Hilbert type matrix with 10^8 lines; in section 5.2 a *big data* linear regression problem with real data. Section 6 is devoted to final remarks and conclusions.

2. The stochastic PLWK method

In what follows we introduce the sPLWK method for solving the linear ill-posed problem (1) and (2). 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 (7) is given (e.g., $x_0 = 0$) as well as a sequence $(\theta_k) \in \mathbb{R}$ of relaxation parameters, and a positive constant γ. 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 $\sup_k \theta_k < 2$;

(A4) We choose $\gamma > C := \max_i \|A_i\|$ (assumption (A2) implies $\max_i \|A_i\| < \infty$);

(A5) The stopping index $k_\delta^* = k^*(\delta)$, satisfies $\lim_{\delta\to 0} k_\delta^* = \infty$ and $\lim_{\delta\to 0} \|\delta\|^2 k_\delta^* = 0$;

(A6) We denote $p_i = \mathbb{P}(I_k = i)$ and assume $p_i \in (0, 1)$, for $i = 0, \ldots, N-1$ (with $\sum_i p_i = 1$).

Here, in a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, (I_k) is an independent and identically distributed sequence of random indexes taking values in $\{0, \ldots, N-1\}$. Notice that, in (A5) the stopping index is a function $k^* : \mathbb{R}^N \ni \delta \to k_\delta^* \in \mathbb{N}$.

2.2. Description of the method

In the sequel we introduce the sPLWK method for solving (1) and (2). Given x_0, (θ_k) and γ as in section 2.1, we consider the sequence $(x_k^\delta) \in X$ generated by the iteration formula

$$x_{k+1}^\delta = x_k^\delta - \theta_k \lambda_{I_k} A_{I_k}^* (A_{I_k} x_k^\delta - y_{I_k}^\delta), \quad k = 0, \ldots, k_\delta^* - 1, \tag{6a}$$

where the stepsize $\lambda_{I_k} := \lambda_{I_k}(x_k^\delta)$ is given by

$$\lambda_{I_k}(x_k^\delta) := \begin{cases} \dfrac{\|A_{I_k}x_k^\delta - y_{I_k}^\delta\| \left(\|A_{I_k}x_k^\delta - y_{I_k}^\delta\| - \delta_{I_k}\right)}{\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\|^2}, & \text{if } \|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| > \gamma\delta_{I_k} \\ 0, & \text{otherwise.} \end{cases} \tag{6b}$$

Observe that, additionally to depending on I_k, $\lambda_{I_k} = \lambda_{I_k}(x_k^\delta)$ also depends on the realization of $I_0, \dots, I_{k-1}$ through the random variable x_k^δ. In particular, given I_k, λ_{I_k} is still random.

The careful reader observes that, differently from deterministic Kaczmarz type methods (e.g., Kaczmarz/ART [20], LWK [14, 16], PLW [24], PLWK [23], LMK [2, 10], EMK [15], iTK [11]), the sPLWK method exhibits no cyclic structure since the choice of the index I_k is independent of the previously chosen indexes I_j for $j = 0, \dots, k-1$.

The stochastic structure of the sPLWK method is motivated by the ideas discussed in [34], 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 $N \gg M$. Aiming to solve (1) and (2) with exact data, the authors propose a non-cyclic method with an iterative step analog to the step of the original Kaczmarz method[6]. It is worth mentioning that, in [34], 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 (Exact projections). The **sPLWK method with exact projections** is obtained by taking $\theta_k = 1$ in (6a), which amounts to define x_{k+1}^δ as the orthogonal projection of x_k^δ onto H_{I_k, x_k^δ}, where

$$H_{i,x} := \left\{z \in X | \langle z - x,\ A_i^*(y_i^\delta - A_i x)\rangle \geqslant \|y_i^\delta - A_i x\| \left(\|y_i^\delta - A_i x\| - \delta_i\right)\right\}$$

(compare with the iterative step of the PLW method in [24]). A relaxed variant of the sPLWK method uses $\theta_k \in (0, 2)$, so that x_{k+1}^δ can be interpreted as a relaxed projection of x_k^δ onto H_{I_k, x_k^δ}.

Remark 2.2 (Separation property). The solution set $A_i^{-1}(y_i)$ of the ith-equation is contained in $H_{i,x}$ for $i = 0, \dots, N-1$ and all $x \in X$. Indeed, for each $x^* \in A_i^{-1}(y_i)$ we have

$$\langle x^* - x, A_i^*(y_i^\delta - A_i x)\rangle = \langle y_i - y_i^\delta + y_i^\delta - A_i x, y_i^\delta - A_i x\rangle \geqslant \|y_i^\delta - A_i x\| \left(\|y_i^\delta - A_i x\| - \delta_i\right).$$

Moreover, from the definition of $H_{i,x}$ follows that $x \in H_{i,x}$ if and only if $\|y_i^\delta - A_i x\| \leqslant \delta_i$. These two facts allow us to conclude that the convex set $H_{i,x}$ **separates** $A_i^{-1}(y_i)$ from $x \in X$ whenever $\|y_i^\delta - A_i x\| > \delta_i$.

Remark 2.3 (Exact data case). Notice that $A_i^*(A_i x - y_i) = 0$ iff $A_i x = y_i$. [7] Therefore, (6b) can be written in the form

$$\lambda_{I_k}(x_k) := \begin{cases} \dfrac{\|A_{I_k}x_k - y_{I_k}\|^2}{\|A_{I_k}^*(A_{I_k}x_k - y_{I_k})\|^2}, & \text{if } \|A_{I_k}x_k - y_{I_k}\| > 0 \\ 0, & \text{otherwise} \end{cases}$$

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

[7] 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, $A_i x - y_i \in R(A_i) \cap R(A_i)^\perp = \{0\}$.

(here (x_k) denotes the sequence generated by (6) using exact data). Consequently, the sPLWK method in (6) can be interpreted as follows:

- If $A_{I_k}x_k \neq y_{I_k}$, then x_{k+1} is given by (6a) with $\lambda_{I_k} = \|A_{I_k}x_k - y_{I_k}\|^2 \|A_{I_k}^*(A_{I_k}x_k - y_{I_k})\|^{-2}$;
- If $A_{I_k}x_k = y_{I_k}$, then $x_{k+1} = x_k$ and $\lambda_{I_k} = 0$.

Notice that $\|y_{I_k} - A_{I_k}x_k\| > 0$ is sufficient to guarantee that the convex set H_{I_k,x_k} separates $A_{I_k}^{-1}(y_{I_k})$ from x_k (see remark 2.2). Thus, for any $\theta_k \in (0,2)$, x_{k+1} given by (6a) is closer to the solution set $A_{I_k}^{-1}(y_{I_k})$ than x_k.

Remark 2.4 (Noisy data case). The sPLWK method in (6) can be interpreted as follows:

- If $\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| > \gamma\delta_{I_k}$, then x_{k+1}^δ is given by (6a) with λ_{I_k} as in (6b);
- If $\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| \leqslant \gamma\delta_{I_k}$, then $x_{k+1}^\delta = x_k^\delta$ and $\lambda_{I_k} = 0$.

Due to Assumption (A4), inequality $\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| > \gamma\delta_{I_k}$ in (6b) implies $\|y_{I_k}^\delta - A_{I_k}x_k^\delta\| > C^{-1}\gamma\delta_{I_k} > \delta_{I_k}$. From remark 2.2 we conclude that, in this case, H_{I_k,x_k^δ} separates $A_{I_k}^{-1}(y_{I_k})$ from x_k^δ. Thus, for any $\theta_k \in (0,2)$, x_{k+1}^δ given by (6a) is closer to the solution set $A_{I_k}^{-1}(y_{I_k})$ than x_k^δ.

Remark 2.5 (Lower bound for the stepsizes λ_{I_k}).

- In the exact data case, assumption (A2) imply $\lambda_{I_k} \geqslant C^{-2}$ whenever $\|A_{I_k}x_k - y_{I_k}\| > 0$ (see also remark 2.2). In other words, C^{-2} is a natural lower bound for the stepsizes defined in (6b), whenever x_k is not a solution of $A_{I_k}x = y_{I_k}$.
- In the noisy data case, assumptions (A2) and (A4) imply $\lambda_{I_k}(x_k^\delta) \geqslant (\gamma - C)(\gamma C^2)^{-1} =: \lambda_{\min}$, whenever $\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| > \gamma\delta_{I_k}$.

3. Exact data case

In this section we analyze the sPLWK method for solving the linear ill-posed problem (1) and (2) in the case of exact data, i.e., $\delta_i = 0$. In this case, the inverse problem can be written in the form

$$A_i x = y_i, \quad i = 0, \ldots, N-1, \tag{7}$$

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

Remark 3.1 (a word on notation). For the remaining of the manuscript we use the notation:

(a) Given $x^* \in X$ a solution of (7), the mean square iteration error $\mathbb{E}[\|x^* - x_k\|^2]$ is defined by the average error over all possible realizations of $I_0, \ldots, I_{k-1}$ that define x_k. In other words, for $k = 0$ and $k = 1$, we find

$$\mathbb{E}\left[\|x^* - x_0\|^2\right] = \|x^* - x_0\|^2, \mathbb{E}\left[\|x^* - x_1\|^2\right] = \sum_{i=0}^{N-1} p_i \ \|x^* - \left[x_0 - \theta_1\lambda_i A_i^*(A_i x_0 - y_i)\right]\|^2.$$

(b) Let $k \in \mathbb{N}$ be fixed. Denote by $\mathcal{F}_k$ the σ-algebra generated by $I_0, \ldots, I_{k-1}$. Then

$$\mathbb{E}\left[\lambda_I \|A_I x_k - y_I\|^2 | \mathcal{F}_k\right] = \sum_{i=0}^{N-1} p_i\, \lambda_i\, \|A_i x_k - y_i\|^2 \quad \text{and}$$

$$\mathbb{E}\left[\|x^* - x_k\|^2 | \mathcal{F}_k\right] = \|x^* - x_k\|^2.$$

Here $\lambda_i = \lambda_i(x_k) := \|A_i x_k - y_i\|^2 \|A_i^*(A_i x_k - y_i)\|^{-2}$, for $i = 0, \ldots, N-1$ (see remark 2.3). I.e., λ_i is a random variable depending on the realization of x_k.

(c) By the law of iterated expectation, we find $\mathbb{E}\left[\|x^* - x_{k+1}\|^2|\right] = \mathbb{E}\left[\mathbb{E}\left[\|x^* - x_{k+1}\|^2 | \mathcal{F}_k\right]\right]$ and

$$\mathbb{E}\left[\lambda_I \|A_I x_k - y_I\|^2\right] = \mathbb{E}\left[\mathbb{E}\left[\lambda_I \|A_I x_k - y_I\|^2 | \mathcal{F}_k\right]\right] = \sum_{i=0}^{N-1} p_i\ \mathbb{E}\left[\lambda_i \|A_i x_k - y_i\|^2\right];$$

the last expectation averages the residual of equation i times λ_i over all possible realizations of x_k.

Moreover, $\mathbb{E}\left[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2\right] = \mathbb{E}\left[\mathbb{E}\left[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2 | \mathcal{F}_k\right]\right]$, where

$$\begin{aligned}
&\mathbb{E}\left[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2 | \mathcal{F}_k\right] \\
&\quad = \sum_{i=0}^{N-1} p_i\ \left(\|x^* - [x_k - \theta_k \lambda_i A_i^*(A_i x_k - y_i)]\|^2 - \|x^* - x_k\|^2\right) \\
&\quad = \sum_{i=0}^{N-1} p_i\, \|x^* - [x_k - \theta_k \lambda_i A_i^*(A_i x_k - y_i)]\|^2 - \sum_{i=0}^{N-1} p_i\, \|x^* - x_k\|^2 \\
&\quad = \mathbb{E}\left[\|x^* - x_{k+1}\|^2 | \mathcal{F}_k\right] - \|x^* - x_k\|^2.
\end{aligned}$$

(d) For each $k \in \mathbb{N}$ it holds, by linearity of expectation,

$$\mathbb{E}\left[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2\right] \ = \ \mathbb{E}\left[\|x^* - x_{k+1}\|^2\right] - \mathbb{E}\left[\|x^* - x_k\|^2\right].$$

Indeed, it follows from (b) $\mathbb{E}\,[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2\,] = \mathbb{E}\,[\mathbb{E}\,[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2 | \mathcal{F}_k\,]\,] = \mathbb{E}\,[\mathbb{E}\,[\|x^* - x_{k+1}\|^2 | \mathcal{F}_k\,] - \mathbb{E}\,[\|x^* - x_k\|^2 | \mathcal{F}_k\,]\,] = \mathbb{E}\,[\mathbb{E}\,[\|x^* - x_{k+1}\|^2 | \mathcal{F}_k\,]\,] - \mathbb{E}\,[\mathbb{E}\,[\|x^* - x_k\|^2 | \mathcal{F}_k\,]\,] = \mathbb{E}\,[\|x^* - x_{k+1}\|^2\,] - \mathbb{E}\,[\|x^* - x_k\|^2\,]$.

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 (7). This is a fundamental result for the forthcoming analysis.

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

$$\mathbb{E}\left[\|x^* - x_{k+1}\|^2\right] - \mathbb{E}\left[\|x^* - x_k\|^2\right] = \theta_k(\theta_k - 2)\,\mathbb{E}\left[\lambda_I \|A_I x_k - y_I\|^2\right], \quad k = 0, 1, \ldots \tag{8}$$

(note that $\lambda_I = \lambda_I(x_k)$ and expectation in (8) should be understood as in (c) of remark 3.1). Moreover, it holds

$$\mathbb{E}\left[\|x^* - x_{k+1}\|^2\right] - \mathbb{E}\left[\|x^* - x_k\|^2\right] \leqslant C^{-2}\theta_k(\theta_k - 2)\,\mathbb{E}\left[\|A_I x_k - y_I\|^2\right], \quad k = 0, 1, \ldots$$

Proof. If $A_{I_k}^*(A_{I_k}x_k - y_{I_k}) \neq 0$, then it follows from (A1) that $\bigcap_i A_i^{-1}(y_i) \neq \emptyset$. Thus, for any x^* solution of (7) we have

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

Otherwise, if $A_{I_k}^*(A_{I_k}x_k - y_{I_k}) = 0$, then $x_{k+1} = x_k$ and $A_{I_k}x_k = y_{I_k}$. Consequently, (9) holds in this case a well.

Denoting by $\mathcal{F}_k$ the σ-algebra generated by $(I_0, \ldots, I_{k-1})$, we notice that x_k is measurable with respect to $\mathcal{F}_k$, and I_k is independent of it. Thus, it follows from (9) that

$$\mathbb{E}\left[\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2|\mathcal{F}_k\right] = \theta_k(\theta_k - 2)\,\mathbb{E}\left[\lambda_I\|A_Ix_k - y_I\|^2|\mathcal{F}_k\right].$$

Now, taking full expectation yields (8) (see remark 3.1 (c) and (d)). To conclude the proof, notice that the second assertion follows from (8) together with remark 2.5. □

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

Proposition 3.3. *Let the assumptions of lemma 3.2 hold. Additionally, let assumption (A3) hold. Then, for any x^* solution of (7) we have*

$$\mathbb{E}[\|x^* - x_{k+1}\|^2] \leqslant \mathbb{E}[\|x^* - x_k\|^2], \quad k = 0, 1, \ldots \qquad (10)$$

Another consequence of lemma 3.2 is discussed in the next proposition. This result is needed for the proof of theorem 3.6 (convergence for exact data).

Proposition 3.4. *Let the assumptions of lemma 3.2 hold. Additionally, let assumption (A3) hold. Then, the series*

$$\sum_{k=0}^{\infty}\theta_k(2-\theta_k)\mathbb{E}[\lambda_I\|A_Ix_k - y_I\|^2]\,, \quad \sum_{k=0}^{\infty}\theta_k\,\mathbb{E}[\lambda_I\|A_Ix_k - y_I\|^2] \quad \textit{and}$$

$$\sum_{k=0}^{\infty}\mathbb{E}[\|A_Ix_k - y_I\|^2]$$

are all summable.

Proof. The summability of the first series follows from lemma 3.2. The summability of the second series follows from (A3) and the summability of the first series. The summability of the third series follows from (A3), the summability of the second series, and the facts:

(a) $\lambda_{I_k} = 0$ iff $A_{I_k}x_k = y_{I_k}$;
(b) $\lambda_{I_k} \geqslant C^{-2}$ whenever $\|A^*_{I_k}(A_{I_k}x_k - y_{I_k})\| > 0$
(see remarks 2.2 and 2.3). □

Yet another consequence of lemma 3.2 is the fact that the sequence (x_k) generated by the rPLKW method with exact projections (i.e., obtained by choosing $\theta_k = 1$ in (6a)) is an **average reasonable wanderer** in the sense of [4], i.e., $\sum_k \mathbb{E}[\|x_k - x_{k+1}\|^2] < \infty$. Indeed, since $\theta_k = 1$, it follows from (6) 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 2.2). 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 $A_{I_k}x = y_{I_k}$.[8] 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 proposition 3.4 we conclude that $\sum_{k=0}^{\infty} \mathbb{E}[\|x_k - x_{k+1}\|^2] < \infty$.

We are now ready to state and prove the main result of this manuscript, namely convergence in mean square of the sPLWK method. First, however, we briefly recall the concept of **minimal norm solutions** of (7).

Remark 3.5. It is worth noticing that there exists an x_0-minimal norm solution of (7), i.e., a solution $x^\dagger$ of (7) satisfying $\|x^\dagger - x_0\| = \inf\{\|x - x_0\|;\ x \in X \ \text{ is solution of (7)}\}$.[9] Moreover, $x^\dagger$ is the only solution of (7) with this property.

Theorem 3.6 (Convergence for exact data). *Let assumptions (A1), (A2), (A3) and (A6) hold true. Then, any sequence (x_k) generated by the sPLWK method (6) converges in mean square to $x^\dagger$, the x_0-minimal norm solution of (7): $\mathbb{E}[\|x^\dagger - x_k\|^2] \to 0$ 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 proposition 3.3 follows

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

for some $\varepsilon \geqslant 0$. 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 $k \leqslant l$ for some $k \leqslant n \leqslant l$ (compare with [17, theorem 2.3]). Notice that $\mathbb{E}[\langle\cdot\,,\cdot\rangle_X]$, $\mathbb{E}[\langle\cdot\,,\cdot\rangle_Y]$) define inner products in $L^2(\Omega; X)$ and $L^2(\Omega; Y)$ respectively[10].

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

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

[8] Notice that all solutions of $A_{I_k}x = y_{I_k}$ belong to H_{I_k,x_k}.

[9] See, e.g., [9] for details.

[10] $L^2(\Omega; X)$ is the space of square integrable random variables defined on Ω and taking values in X. We write $\mathbb{E}[\langle\cdot\,,\cdot\rangle]$ instead of $\mathbb{E}[\langle\cdot\,,\cdot\rangle_X]$ whenever it is clear from the context.

holds true. Next, we argue with (6a) and the Cauchy–Schwartz inequality to estimate

$$\begin{aligned}
|\mathbb{E}[\langle e_n - e_k, e_n\rangle]| &= \left|\sum_{j=k}^{n-1} \mathbb{E}[\langle x_{j+1} - x_j, x^\star - x_n\rangle]\right| \\
&= \left|\sum_{j=k}^{n-1} \mathbb{E}[\theta_j\, \lambda_I \langle A_I^*(y_I - A_I x_j), x^\star - x_n\rangle]\right| \\
&= \left|\sum_{j=k}^{n-1} \theta_j\, \mathbb{E}[\lambda_I \langle y_I - A_I x_j, A_I(x^\star - x_n)\rangle]\right| \\
&= \left|\sum_{j=k}^{n-1} \theta_j\, \mathbb{E}\left[\langle \lambda_I^{\frac{1}{2}}(y_I - A_I x_j), \lambda_I^{\frac{1}{2}}(y_I - A_I x_n)\rangle\right]\right| \\
&\leqslant \sum_{j=k}^{n-1} \theta_j\, \mathbb{E}[\lambda_I \|A_I x_j - y_I\|^2]^{\frac{1}{2}}\ \mathbb{E}[\lambda_I \|A_I x_n - y_I\|^2]^{\frac{1}{2}}. \qquad (14)
\end{aligned}$$

Now notice that, due to minimizing property (13), it follows from (14)

$$|\mathbb{E}[\langle e_n - e_k, e_n\rangle]| \ \leqslant\ \sum_{j=k}^{n-1} \theta_j\, \mathbb{E}[\lambda_I \|A_I x_j - y_I\|^2].$$

Consequently, proposition 3.4 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 $\mathbb{E}[\|e_j - e_k\|^2]^{\frac{1}{2}} \leqslant \mathbb{E}[\|e_j - e_l\|^2]^{\frac{1}{2}} + \mathbb{E}[\|e_l - e_k\|^2]^{\frac{1}{2}}$ 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 (7).

Since (x_k) is Cauchy in $L^2(\Omega; X)$, it has an accumulation point x^*. Moreover, it follows from proposition 3.4 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 $i = 0, \ldots, N-1$ (at this point assumption (A6) is needed). Thus x^* is a solution of (7).

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 $k = 0, 1, \ldots$.[11] Thus, an inductive argument shows that $x^* \in x_0 + \mathcal{N}(\mathbf{A})^\perp$. However, $x^\dagger$ is the only solution of (7) with this property (see remark 3.5), concluding the proof. □

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

4. Regularization properties

Is this section we investigate the regularization properties of sPLWK method in the noisy data case. The mean square iteration error $\mathbb{E}[\|x^* - x_k^\delta\|^2]$ is defined as in remark 3.1.

In what follows we estimate the *average gain* $\mathbb{E}[\|x^* - x_{k+1}^\delta\|^2] - \mathbb{E}[\|x^* - x_k^\delta\|^2]$, extending the results in lemma 3.2 to the noisy data case.

Lemma 4.1. *Let assumptions* (*A1*) *and* (*A2*) *hold true and* (x_k^δ) *be a sequence generated by the sPLWK method* (6). *Then, for any* x^* *solution of* (7) *it holds*

$$\begin{aligned} &\mathbb{E}\left[\|x^* - x_{k+1}^\delta\|^2\right] - \mathbb{E}\left[\|x^* - x_k^\delta\|^2\right] \\ &\qquad \leqslant \theta_k(\theta_k - 2)\mathbb{E}\left[\lambda_I \|A_I x_k^\delta - y_I^\delta\| \left(\|A_I x_k^\delta - y_I^\delta\| - \delta_I\right)\right], \quad k = 0, \ldots, k_\delta^* - 1. \end{aligned} \tag{15}$$

Proof. Let $I_k \in \{0, \ldots, N-1\}$. If $\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| > \gamma\delta_{I_k}$, we derive from (6)

$$\begin{aligned} &\|x^* - x_{k+1}^\delta\|^2 - \|x^* - x_k^\delta\|^2 \\ &\quad = 2\theta_k\lambda_{I_k}\langle y_{I_k} \pm y_{I_k}^\delta - A_{I_k}x_k^\delta, A_{I_k}x_k^\delta - y_{I_k}^\delta\rangle + \|x_{k+1}^\delta - x_k^\delta\|^2 \\ &\quad \leqslant 2\theta_k\lambda_{I_k}\|A_{I_k}x_k^\delta - y_{I_k}^\delta\| \left(\delta_{I_k} - \|A_{I_k}x_k^\delta - y_{I_k}^\delta\|\right) + \|x_{k+1}^\delta - x_k^\delta\|^2 \\ &\quad = \theta_k(\theta_k - 2)\lambda_{I_k}\|A_{I_k}x_k^\delta - y_{I_k}^\delta\| \left(\|A_{I_k}x_k^\delta - y_{I_k}^\delta\| - \delta_{I_k}\right). \end{aligned} \tag{16}$$

Otherwise, if $\|A_{I_k}^*(A_{I_k}x_k^\delta - y_{I_k}^\delta)\| \leqslant \gamma\delta_{I_k}$, we have $\lambda_{I_k} = 0$, $x_{k+1}^\delta = x_k^\delta$, and (16) holds trivially. The remaining of the proof follows the lines of the proof of lemma 3.2. □

Notice that, due to (6b), the term $\lambda_I\|A_I x_k^\delta - y_I^\delta\|\left(\|A_I x_k^\delta - y_I^\delta\| - \delta_I\right)$ on the right-hand side of (15) is either positive or zero. Consequently, we derive

Proposition 4.2 (Monotonicity). *Let the assumptions of lemma* 4.1 *hold. Additionally, let assumption* (*A3*) *hold. Then*

$$\mathbb{E}\left[\|x^* - x_{k+1}^\delta\|^2\right] \leqslant \mathbb{E}\left[\|x^* - x_k^\delta\|^2\right], \quad k = 0, \ldots, k_\delta^* - 1$$

for any x^* *solution of* (7).

The careful reader observes that, due to the definition of λ_{I_k} in (6b), both lemma 4.1 and proposition 4.2 hold true also for $k > k_\delta^*$ (if one continues to iterate after step k_δ^*).

Theorem 4.3 (Stability). *Let assumptions* (*A1*), (*A2*), (*A3*) *and* (*A6*) *hold. Let* $(\delta^j) = (\delta_0^j, \ldots, \delta_{N-1}^j) \in (\mathbb{R}^+)^N$ *be a sequence with* $\|\delta^j\| \to 0$ *as* $j \to \infty$, *and* $(y^{\delta^j}) = (y_0^{\delta^j}, \ldots, y_{N-1}^{\delta^j}) \in Y^N$ *be a corresponding sequence of noisy data satisfying* (1). *Moreover, let* $(x_l)_{l\in\mathbb{N}}$ *and* $(x_l^{\delta^j})_{l=0}^{k_\delta^*}$ *be the sequences generated by the sPLWK method in the case of exact and noisy data respectively; all sequences are generated using the same* $(I_0, \ldots, I_k, \ldots)$. *Then, for each* $k \in \mathbb{N}$ *it holds*

$$\lim_{j\to\infty}\mathbb{E}\left[\|x_k^{\delta^j} - x_k\|^2\right] = 0. \tag{17}$$

Proof. We give here an inductive proof in k. Notice that $x_0 = x_0^{\delta^j}$, for all $j \in \mathbb{N}$. Consequently, (17) holds true for $k = 0$. Next, assume that $\lim_j \mathbb{E}\left[\|x_k^{\delta^j} - x_k\|^2\right] = 0$. Our goal is to prove that $\lim_j \mathbb{E}\left[\|x_{k+1}^{\delta^j} - x_{k+1}\|^2\right] = 0$. The proof is divided in three steps:

Step 1. We verify that, for each fixed $k \in \mathbb{N}$, $\lim_j \|x_k^{\delta^j} - x_k\|^2 = 0$.

We claim that, for each realization $(I_0, \dots, I_{k-1})$, the inequality $\|x_k^\delta - x_k\|^2 \leqslant c\,\mathbb{E}\left[\|x_k^\delta - x_k\|^2\right]$ holds, where $c = c(k, N, p_0, \dots, p_{N-1})$ is a positive constant. Indeed, for $k = 1$, we have $\mathbb{E}\left[\|x_1^\delta - x_1\|^2\right] = \sum_{i=0}^{N-1} p_i \|x_{1;i}^\delta - x_{1;i}\|^2 \geqslant p_{\min} \sum_{i=0}^{N-1} \|x_{1;i}^\delta - x_{1;i}\|^2 \geqslant p_{\min}\, \|x_{1;I_0}^\delta - x_{1;I_0}\|^2 = p_{\min}\, \|x_1^\delta - x_1\|^2$. Here $p_{\min} := \min_i p_i > 0$ and $x_{1;i}^\delta$ is given by (6a) taking $I_0 = i$ ($x_{1;i}$ is defined analogously). Thus, for $k = 1$ our claim holds with $c = p_{\min}^{-1}$. For $k > 1$ we have

$$\begin{aligned}
\mathbb{E}\left[\|x_k^\delta - x_k\|^2\right] &= \sum_{\substack{i_0=0\\ \cdots\\ i_{k-1}=0}}^{N-1} p_{i_0} \dots p_{i_{k-1}} \|x_{k;i_0,\dots,i_{k-1}}^\delta - x_{k;i_0,\dots,i_{k-1}}\|^2 \\
&\geqslant p_{\min}^k \|x_{k;I_0,\dots,I_{k-1}}^\delta - x_{k;I_0,\dots,I_{k-1}}\|^2 = p_{\min}^k \|x_k^\delta - x_k\|^2,
\end{aligned}$$

where $x_{k;i_0,\dots,i_{k-1}}^\delta$ is given by (6a) taking $(I_0, \dots, I_{k-1}) = (i_0, \dots, i_{k-1})$ ($x_{k;i_0,\dots,i_{k-1}}$ is defined analogously). Consequently, our claim holds with $c = p_{\min}^{-k}$.

Step 1 follows now from the inductive hypothesis, the fact that $x_k^{\delta^j}$ and x_k are both generated by the same fixed $(I_0, \dots, I_{k-1})$, and the above claim with $\delta = \delta^j$.

Step 2. We verify that, for each fixed $k \in \mathbb{N}$, $\lim_j \|x_{k+1}^{\delta^j} - x_{k+1}\|^2 = 0$.

Step 2(a). The case $A_{I_k}^*(A_{I_k} x_k - y_{I_k}) \neq 0$.

It follows from (1), (A2) and step 1 that $\lim_j \|A_{I_k}^*(A_{I_k} x_k^{\delta^j} - y_{I_k}^{\delta^j})\| = \|A_{I_k}^*(A_{I_k} x_k - y_{I_k})\| > 0$. Thus, we have $\|A_{I_k}^*(A_{I_k} x_k^{\delta^j} - y_{I_k}^{\delta^j})\| > \frac{1}{2}\|A_{I_k}^*(A_{I_k} x_k - y_{I_k})\| > \gamma \delta_{I_k}^j$ for sufficiently large j (this follows from the previous limit and from the fact that $\lim_j \delta_{I_k}^j = 0$). Consequently, it follows from (6b) that

$$\lim_{j\to\infty} |\lambda_{I_k}^{\delta^j} - \lambda_{I_k}| = 0 \tag{18}$$

(here we distinguish λ_{I_k} from $\lambda_{I_k}^\delta$ for exact and noisy data, respectively). In particular, the sequence $(\lambda_{I_k}^{\delta^j})_j$ is bounded.

Moreover, it follows from the iteration formula (6a)

$$\begin{aligned}
&x_{k+1}^{\delta^j} - x_{k+1} \\
&\quad = x_k^{\delta^j} - x_k - \theta_k \left[\lambda_{I_k}^{\delta^j} A_{I_k}^*(A_{I_k} x_k^{\delta^j} - y_{I_k}^{\delta^j}) - \lambda_{I_k} A_{I_k}^*(A_{I_k} x_k - y_{I_k})\right] \\
&\quad = x_k^{\delta^j} - x_k - \theta_k(\lambda_{I_k}^{\delta^j} - \lambda_{I_k}) A_{I_k}^*(A_{I_k} x_k - y_{I_k}) \\
&\qquad - \theta_k \lambda_{I_k}^{\delta^j} A_{I_k}^* \left[(A_{I_k} x_k^{\delta^j} - y_{I_k}^{\delta^j}) - (A_{I_k} x_k - y_{I_k})\right],
\end{aligned}$$

from where we estimate

$$\begin{aligned}
\|x_{k+1}^{\delta^j} - x_{k+1}\| \leqslant{}& \|x_k^{\delta^j} - x_k\| + 2C|\lambda_{I_k}^{\delta^j} - \lambda_{I_k}|\; \|A_{I_k} x_k - y_{I_k}\| \\
&+ 2C\lambda_{I_k}^{\delta^j}\left[C\|x_k^{\delta^j} - x_k\| + \delta_{I_k}^j\right].
\end{aligned}$$

Therefore, it follows from step 1 and (18) that $\lim_j \|x_{k+1}^{\delta^j} - x_{k+1}\|^2 = 0$.

Step 2(b). The case $A_{I_k}^*(A_{I_k}x_k - y_{I_k}) = 0$.

In this case $\lambda_{I_k} = 0$ and $A_{I_k}x_k = y_{I_k}$.[7] Thus, $x_{k+1}^{\delta j} - x_{k+1} = x_k^{\delta j} - x_k - \theta_k \lambda_{I_k}^{\delta j} A_{I_k}^* (A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})$ and we conclude that

$$\|x_{k+1}^{\delta j} - x_{k+1}\| \leqslant \|x_k^{\delta j} - x_k\| + 2\lambda_{I_k}^{\delta j} \|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\|. \tag{19}$$

If $\|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\| \leqslant \gamma\delta_{I_k}^j$ then $\lambda_{I_k}^{\delta j} = 0$ and it follows from (19)

$$\|x_{k+1}^{\delta j} - x_{k+1}\| \leqslant \|x_k^{\delta j} - x_k\|. \tag{20}$$

Otherwise, if $\|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\| > \gamma\delta_{I_k}^j$, we estimate

$$\begin{aligned}\|A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}\|^2 &= \langle A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}, A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}\rangle \\ &= \langle A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}), x_k^{\delta j} - x_k\rangle \\ &\quad + \langle A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}, A_{I_k}x_k - y_{I_k} + y_{I_k} - y_{I_k}^{\delta j}\rangle \\ &\leqslant \|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\| \|x_k^{\delta j} - x_k\| + \delta^j \|A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}\|.\end{aligned}$$

This inequality and (6b) yields the estimate $\lambda_{I_k}^{\delta j}\|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\| \leqslant \|x_k^{\delta j} - x_k\|$. From the last inequality and (19) follows

$$\|x_{k+1}^{\delta j} - x_{k+1}\| \leqslant 3\|x_k^{\delta j} - x_k\|. \tag{21}$$

Arguing with step 1, (20) and (21) it follows that, in either case, $\lim_j \|x_{k+1}^{\delta j} - x_{k+1}\|^2 = 0$.
Step 3. We verify that, for each fixed $k \in \mathbb{N}$, $\lim_j \mathbb{E}\left[\|x_{k+1}^{\delta j} - x_{k+1}\|^2\right] = 0$.

Taking the average in step 2 over all possible realizations $(I_0, \ldots, I_k)$ and using the fact that

$$\mathbb{E}\left[\|x_{k+1}^{\delta} - x_{k+1}\|^2\right] = \sum_{\substack{i_0=0\\ \cdots\\ i_k=0}}^{N} p_{i_0}\ldots p_{i_k}\|x_{k;i_0,\ldots,i_k}^{\delta} - x_{k;i_0,\ldots,i_k}\|^2,$$

one concludes that $\lim_j \mathbb{E}\left[\|x_{k+1}^{\delta j} - x_{k+1}\|^2\right] = 0$, completing the inductive proof. □

Notice that, in step 1 of the above proof, the constant c is given by $c = c(k, N, p_0, \ldots, p_{N-1}) = p_{\min}^{-k}$ and becomes unbounded as $k \to \infty$. This fact does not interfere with the result, since k is a fixed (but arbitrary) positive integer.

Theorem 4.4 (Semi-convergence). *Let assumptions (A1), ..., (A6) hold. Let* $(\delta^j) = (\delta_0^j, \ldots, \delta_{N-1}^j) \in \mathbb{R}^N$ *be a zero sequence,* $(\mathbf{y}^{\delta^j}) = (y_0^{\delta^j}, \ldots, y_{N-1}^{\delta^j}) \in Y^N$ *a corresponding sequence of noisy data satisfying* (1). *Moreover, for each* $j \in \mathbb{N}$, *let* $(x_k^{\delta j})_{k=0}^{k^*(\delta^j)}$ *be the corresponding sequence generated by the sPLWK method (these sequences are generated using the same* $(I_0, \ldots, I_k, \ldots)$*). Then we have*

$$\lim_{j\to\infty} \mathbb{E}\left[\|x_{k^*(\delta^j)}^{\delta j} - x^\dagger\|^2\right] = 0, \tag{22}$$

were $x^\dagger$ *is the* x_0*-minimal norm solution of* (7) *(see remark* 3.5*).*

Proof. We claim that

$$\|x^\dagger - x_{k+1}^{\delta j}\|^2 - \|x^\dagger - x_k^{\delta j}\|^2 < \lambda_{\min}^2\,\gamma^2(\delta_{\min}^j)^2 \tag{23}$$

(here $\delta_{\min}^j = \min_{i\in\{0,\ldots,N-1\}} \delta_i^j > 0$). Indeed, if $\|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\| \leqslant \gamma\delta_{I_k}^j$, then $\lambda_{I_k}^{\delta j} = 0$, $x_{k+1}^{\delta j} = x_k^{\delta j}$ and (23) holds trivially. Otherwise, it follows from (16) and remark 2.5

$$\begin{aligned}\|x^\dagger - x_k^{\delta j}\|^2 - \|x^\dagger - x_{k+1}^{\delta j}\|^2 &\geqslant \theta_k(2-\theta_k)\,\lambda_{I_k}\|A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}\|\left(\|A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j}\| - \delta_{I_k}^j\right)\\ &= \theta_k(2-\theta_k)\,\lambda_{I_k}^2\|A_{I_k}^*(A_{I_k}x_k^{\delta j} - y_{I_k}^{\delta j})\|^2\\ &> \theta_k(2-\theta_k)\,\lambda_{\min}^2(\gamma\,\delta_{I_k}^j)^2\\ &\geqslant \theta_k(2-\theta_k)\lambda_{\min}^2\gamma^2\,(\delta_{\min}^j)^2\end{aligned}$$

(notice that, due to (A3), it holds $\theta_k(2-\theta_k) > 0$, for $k = 0, 1, \ldots$). Consequently,

$$\|x^\dagger - x_{k+1}^{\delta j}\|^2 - \|x^\dagger - x_k^{\delta j}\|^2 < |\theta_k(2-\theta_k)|\lambda_{\min}^2\gamma^2(\delta_{\min}^j)^2 < \lambda_{\min}^2\gamma^2(\delta_{\min}^j)^2$$

(notice that from (A3) follows $|\theta_k(2-\theta_k)| < 1$, for $k = 0, 1, \ldots$), which is exactly what we claim in (23).

Now, taking the average in (23) over all possible realizations $(I_0, \ldots, I_k)$ and using remark 3.1 (d) we obtain

$$\mathbb{E}\left[\|x^\dagger - x_{k+1}^{\delta j}\|^2\right] - \mathbb{E}\left[\|x^\dagger - x_k^{\delta j}\|^2\right] < \lambda_{\min}^2\gamma^2(\delta_{\min}^j)^2.$$

Due to (A5) we may assume that $k_{\delta^j}^* = k^*(\delta^j)$ increases strictly monotonically with j. Given $m < n$, we add the inequality above, with $j = n$, from $k = k_{\delta^m}^*$ to $k_{\delta^n}^* - 1$, with the simplified notation $k_j^* = k_{\delta^j}^*$, to obtain

$$\begin{aligned}\mathbb{E}\left[\|x^\dagger - x_{k_n^*}^{\delta^n}\|^2\right] &\leqslant \mathbb{E}\left[\|x^\dagger - x_{k_m^*}^{\delta^n}\|^2\right] + \sum_{k=k_m^*}^{k_n^*-1}\lambda_{\min}^2\gamma^2(\delta_{\min}^n)^2\\ &\leqslant 2\mathbb{E}\left[\|x^\dagger - x_{k_m^*}\|^2\right] + 2\mathbb{E}\left[\|x_{k_m^*} - x_{k_m^*}^{\delta^n}\|^2\right] + \lambda_{\min}^2\gamma^2(\delta_{\min}^n)^2\sum_{k=k_m^*}^{k_n^*-1} 1\\ &\leqslant 2\mathbb{E}\left[\|x^\dagger - x_{k_m^*}\|^2\right] + 2\mathbb{E}\left[\|x_{k_m^*} - x_{k_m^*}^{\delta^n}\|^2\right] + \lambda_{\min}^2\gamma^2\,\|\delta^n\|^2\,k_n^*.\end{aligned}$$

Here (x_k) denotes the sequence generated by (6) using exact data and the same $(I_0, \ldots, I_k, \ldots)$ as the sequences $(x_k^{\delta j})$.

Theorem 3.6 guarantees the existence of a large enough m, s.t. the first term $2\mathbb{E}\left[\|x^\dagger - x_{k_m^*}\|^2\right]$ is smaller then $\varepsilon/3$. Next, from theorem 4.3 with $k = k_m^*$ we conclude that the second term $2\mathbb{E}\left[\|x_{k_m^*} - x_{k_m^*}^{\delta^n}\|^2\right]$ is smaller than $\varepsilon/3$ for large enough n. Finally, due to assumption (A5), the last term $\lambda_{\min}^2\gamma^2\|\delta^n\|^2 k_n^*$ also becomes smaller than $\varepsilon/3$ for large enough n, concluding the proof. □

Table 1. Descriptive Statistics of the probabilities $(p_i)_{i=0}^{N-1}$ for both inverse problems considered in section 5. The first two rows are the mean and the standard deviation, respectively. The last five rows are the quantiles (25%, 50%, 75%), the minimum and the maximum.

	Benchmark	Big Data
Mean	1.0000×10^{-6}	2.3876×10^{-7}
Std	3.2346×10^{-4}	1.0685×10^{-7}
min	1.1157×10^{-11}	6.2049×10^{-8}
25%	1.9834×10^{-11}	1.3341×10^{-7}
Median	4.4624×10^{-11}	2.4929×10^{-7}
75%	1.7847×10^{-10}	3.2199×10^{-7}
Max	2.8407×10^{-1}	4.6939×10^{-7}

5. Numerical experiments

5.1. A benchmark problem

In this section the sPLWK method in (6) is implemented for solving a benchmark problem, which happens to be a well-known system of linear ill-posed equations[12].

Let $\mathbf{H} = (H_i)_{i=0}^{N-1} \in \mathbb{R}^{N\times M}$ be a Hilbert type matrix with rows $H_i = \left(\frac{1}{i+j+1}\right)_{j=0}^{M-1} \in \mathbb{R}^{1,M}$, $X = \mathbb{R}^{M,1}$ and $Y = \mathbb{R}$, where $N = 10^8$ and $M = 2^6$.

The operator $\mathbf{A} = (A_i)_{i=0}^{N-1} \in \mathbb{R}^{N\times M}$, with rows $A_i \in \mathbb{R}^{1,M}$, is obtained by a random shuffle of the rows of $\mathbf{H}$. In our numerical experiments we set $x^\star = (1, \ldots, 1) \in X$ and compute the corresponding exact data $y_i = A_i x^\star$. Two distinct datasets $y^{\delta_1}, y^{\delta_2} \in Y^N$ corresponding to distinct levels of noise are used: (1) $\delta_1 = 10^{-16}$, what corresponds to the MATLAB double precision accuracy; (2) $\delta_2 = 10^{-1}$. The performance of the sPLWK method is compared against two concurrent randomized Kaczmarz type methods, namely: (1) Landweber–Kaczmarz (LWK) with random ordering of equations within cycles [24]; (2) PLWK with random ordering of equations within cycles [23].

In order to better investigate the behavior of iteration (6), six different runs of the sPLWK method are computed for each set of data. In the first five runs (run 1 to run 5), 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 *) however, 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 [34] for the randomized Kaczmarz iteration). Some descriptive statistics of the probabilities are shown in table 1. Moreover, in figure 5 we show plots of these probabilities.

In figures 1 and 2 we present the numerical results obtained for the datasets (y^{δ_1}, δ_1) and (y^{δ_2}, δ_2) respectively. In both noise scenarios, the first five runs of the sPLWK method (run 1 to run 5) produced similar results. The last run of the sPLWK method (run *) delivered the best numerical performance. This difference is most probably explained by the fact that most of the lines A_i have norms close to $\min_i \|A_i\|$ (see the (TOP) picture in figure 5). Consequently only a few probabilities p_i are high (while the others are very small) and the sPLWK method uses, with extremely high frequency, these lines along the iteration.

In figures 1 and 2, the first plot (TOP) shows the evolution of the **residual**, while the second plot (BOTTOM) shows the **relative iteration error**. For the dataset (y^{δ_1}, δ_1) all methods are

[12] Computations are performed using MATLAB® R2017a, running in a Intel® Core ™ i9-10900 CPU.

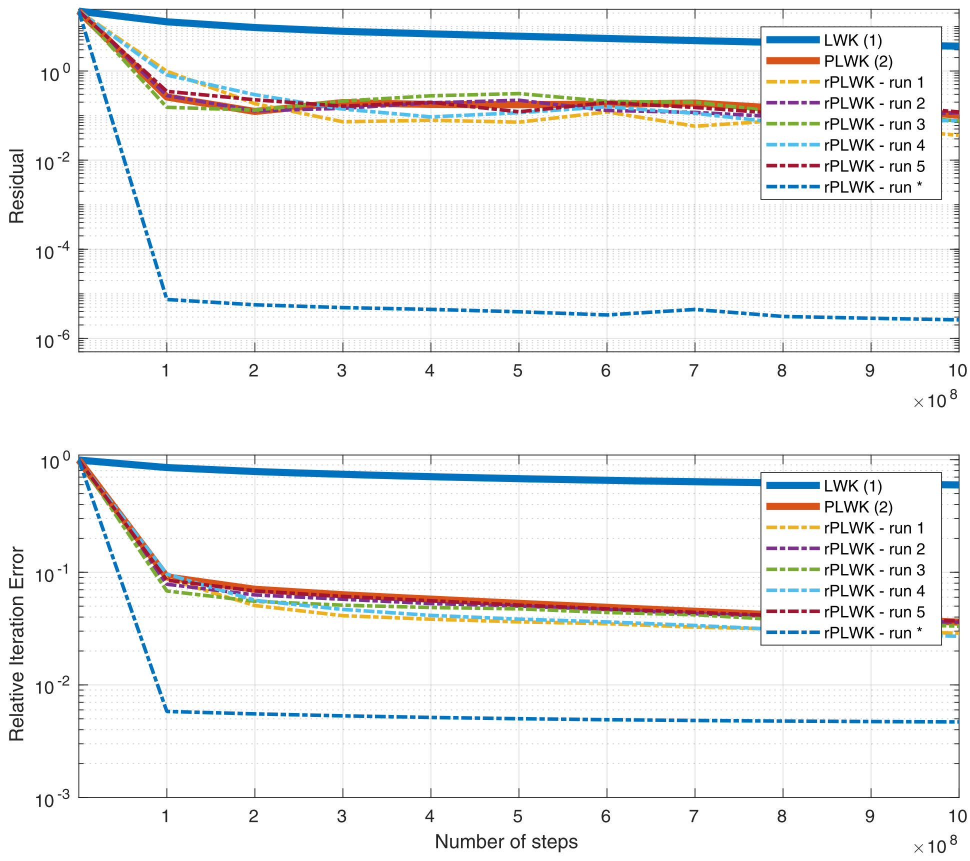

Figure 1. Benchmark problem—dataset (y^{δ_1}, δ_1): (TOP) $\|\mathbf{A}x_k - y\|$; (BOTTOM) $\|x^\star - x_k\|/\|x^\star\|$.

stopped after $10N$ iterative steps (i.e., 10 cycles of methods (1) and (2)), while for (y^{δ_2}, δ_2) the methods are stopped after $5N$ iterations. In both noise scenarios, the sPLWK methods deliver their best results already after N steps.

It is worth noticing that, in the sPLWK method we have monotonicity of the mean square iteration error $\mathbb{E}\left[\|x^\star - x_k\|^2\right]$, see proposition 3.3. However monotonicity of mean square residual $\mathbb{E}\left[\|\mathbf{A}x_k - \mathbf{y}\|^2\right]$ cannot be guaranteed. These two facts are illustrated in both figures 1 and 2.

Remark 5.1 (Methods used in the comparison of numerical performance). The performance of the sPLWK method is compared against two randomized (cyclic) Kaczmarz type methods, namely: LWK [24] and PLWK [23].

It has been observed by many authors (including ourselves [23]) that, numerically, randomized (cyclic) Kaczmarz methods eventually perform better than standard Kaczmarz methods. In the applications considered in sections 5.1 and 5.2, this is not the case. Namely, 'randomized (cyclic) Kaczmarz' and 'standard Kaczmarz' methods perform similarly for the specific inverse problems under consideration. For this reason, no standard Kaczmarz methods are included in section 5 in the comparison of numerical performance.

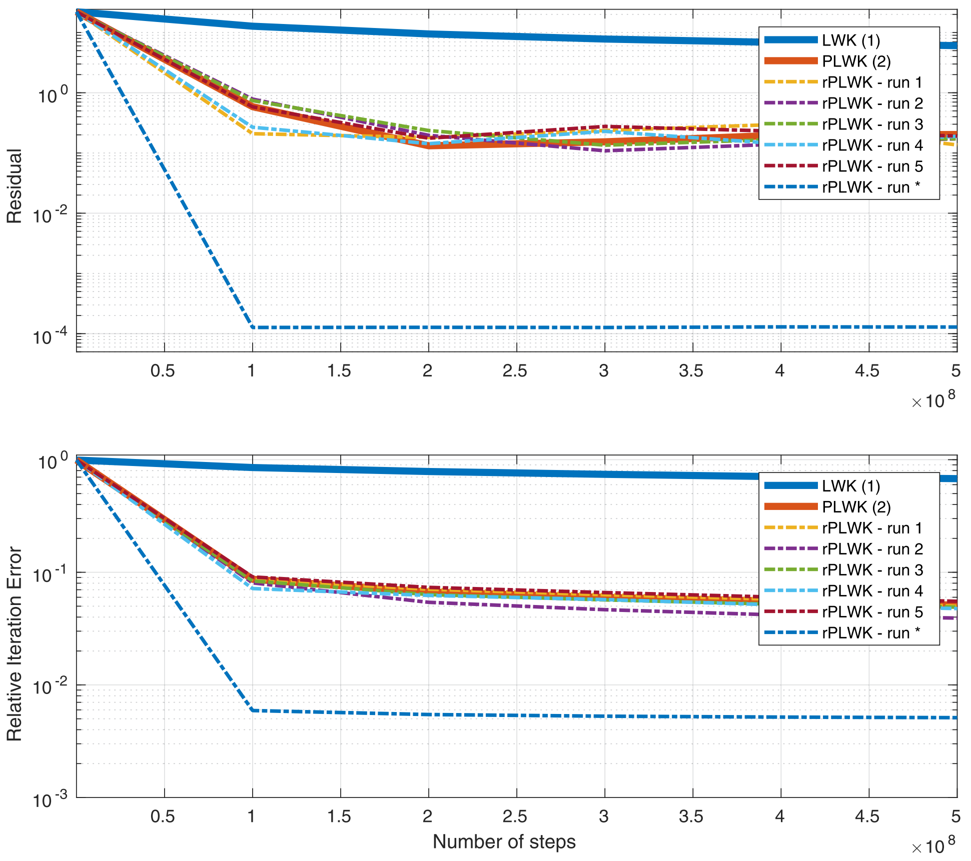

Figure 2. Benchmark problem—dataset (y^{δ_2}, δ_2): (TOP) $\|\mathbf{A}x_k^\delta - y^\delta\|$; (BOTTOM) $\|x^\star - x_k^\delta\|/\|x^\star\|$.

5.2. A big data linear regression problem with real data

Multiple linear regression is a well-known statistical tool for modeling linear relationship between two data sets. More precisely, given n observations of a set of *independent variables* $a_1, \ldots, a_p \in \mathbb{R}^n$, and of a *dependent variable* $y \in \mathbb{R}^n$, one seeks the *regression coefficients* $x_i \in \mathbb{R}$, $i = 0, \ldots, p$ in the linear model $x_0 + a_1 x_1 + \cdots a_p x_p = y$. The corresponding mathematical model can be expressed as $Ax = y$, where $y \in \mathbb{R}^n$ is a vector of n observations of the dependent variable, the columns of $A \in \mathbb{R}^{n\times(p+1)}$ are *predictors* (or *covariates*) and the vector $x \in \mathbb{R}^{p+1}$ contains the regression coefficients. One aims to solve the unconstrained least square problem: $\min_x \|Ax - y\|^2$.

For the specific application considered in this section, we use a data set collected in a gas delivery platform facility at the ChemoSignals Laboratory in the BioCircuits Institute at University of California, San Diego. This measurement system platform provides versatility for obtaining the desired concentrations of the chemical substances of interest with high accuracy.

The data set contains the readings of 16 chemical sensors (Figaro Inc., US) of 4 different types: TGS-2600, TGS-2602, TGS-2610, TGS-2620 (4 units of each type). These sensors

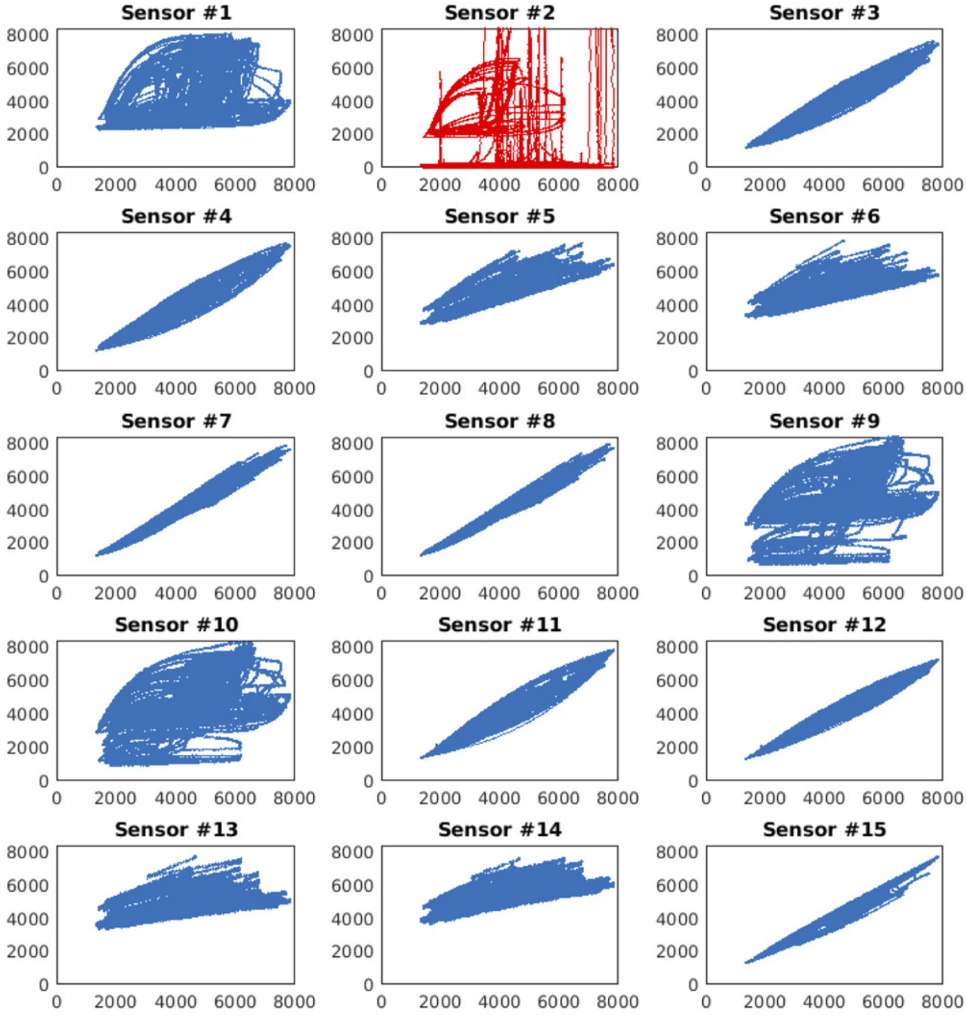

Figure 3. Scatter plots of sensor #i data (for $i = 1, \ldots, 15$) against sensor #16 data.

were exposed to the *mixture of Ethylene and CO* at varying concentrations in air (Ethylene concentration ranges from 0–20 ppm, while CO concentration ranges from 0–600 ppm). For this gas mixture, the measurement was constructed by the continuous acquisition of the 16-sensor array signals for a duration of approximately 12 h without interruption. Concentration transitions were set at random times, in the interval 80–120 s, and to random concentration levels. The data set was constructed such that all possible transitions are present: increasing, decreasing, or setting to zero the concentration of one volatile while the concentration of the other volatile is kept constant (either at a fixed or at zero concentration level). At the beginning, ending, and approximately every 10^4 s, additional predefined concentration patterns with pure gas mixtures were inserted. The concentration ranges for Ethylene and CO were selected such that the induced magnitudes of the sensor responses were similar. Moreover, for gas mixtures, lower concentration levels were favored. Therefore, the multivariate response of the sensors

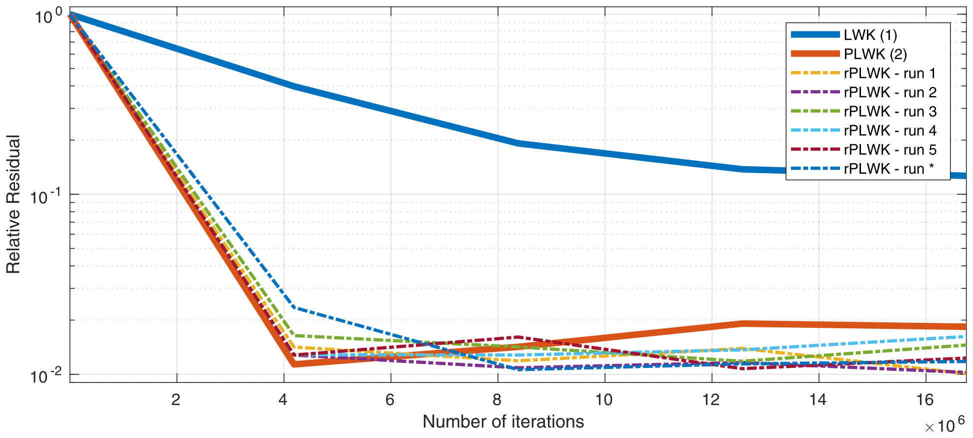

Figure 4. Big data problem: relative residual $\|Ax_k^\delta - y^\delta\| / \|Ax_0 - y^\delta\|$.

to the presented set of stimuli is challenging since the gas mixture configuration can be easily identified from the magnitude of sensors responses (we refer to [12] for further details)[13].

In our numerical tests we follow the experimental setting proposed in [38]: readings from the last sensor (sensor #16) are used as the response variable, and readings from other sensors as covariates (the readings from sensor #2 are not used in the analysis, since approximately 50% of the values are negative). Thus, there are $p = 14$ covariates in this example. Moreover, the first 20 000 data points of all sensors are excluded (this corresponds to less than 4 min of system run-in time). Consequently, each sensor data consists of $n = 4188\,262$ scalar measurements. In figure 3 the scatter plots of sensor #i data against sensor #16 data are presented for $i = 1, \ldots, 15$ (full data is plotted).

Summarizing, the inverse problem under consideration consists in finding approximate solutions to a linear system $\mathbf{A}x = y^\delta$, where $\mathbf{A} = (A_i)_{i=0}^{N-1} \in \mathbb{R}^{N\times M}$, $N = 4188\,262$, $M = 15$ and unknown noise level $\delta > 0$. As in section 5.1, the performance of the sPLWK method is compared against the concurrent Kaczmarz type methods LWK and PLWK. Once again, six different runs of the sPLWK method are computed (these runs are executed as described in section 5.1). All iterations are started using $x_0 = 0$.

In figure 4 we present the evolution of the **relative residual** $\|Ax_k^\delta - y^\delta\| / \|Ax_0 - y^\delta\|$ obtained for the above described experimental setting; all methods are stopped after $4N$ iterative steps. The approximate solution computed by run 2 after N steps reads

$$x_N^\delta = (0.03\,0.00 - 0.79\,0.00\,0.03\,0.29\,0.06\,0.72\,0.040.03\,0.04\,1.12 - 0.39\,0.10 - 0.15)\,.$$

This approximate solution is in agreement with the data plotted in figure 3. Indeed, notice that coordinates $x_{N,i}^\delta$ for $i = 1, 8, 9$ (highlighted red) correspond to sensors #1, #9 and #10 respectively. These coordinates of x_N^δ have small absolute values, while the corresponding sensors have the highest scatter patterns against sensor #16.

The careful reader observes that all runs of the sPLWK method produced similar results. This scenario differs from one observed in section 5.1. A possible explanation resides in the fact that the probabilities p_i used in run * (which are proportional to $\|A_i\|^2$) are similar to

[13] The data set used in our numerical experiments is freely accessible on-line, at the web-site of the UC Irvine Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php).

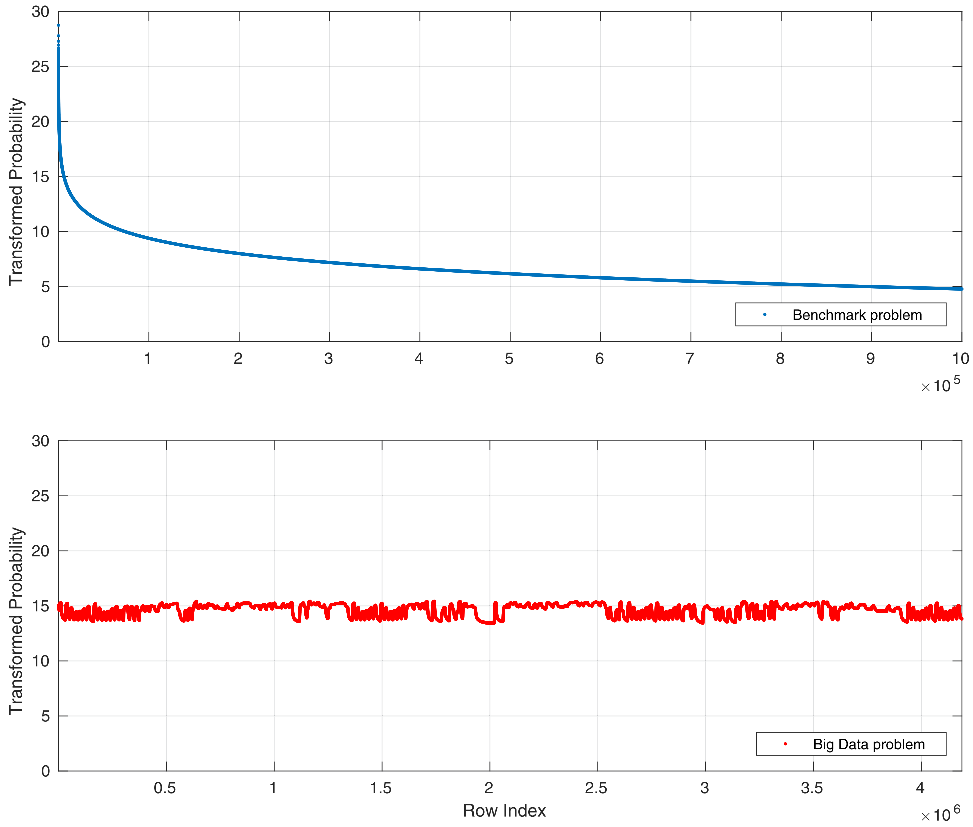

Figure 5. Plots of the probabilities $(p_i)_{i=0}^{N-1}$ for both inverse problems considered in section 5. The probabilities were transformed for a better visualization; we show plots of $30 + \log p_i$, $i = 0, \ldots, N-1$ (see table 1 for details). (TOP) Benchmark problem; (BOTTOM) big data problem.

the uniform probabilities $p_i = N^{-1} = 2.3876 \times 10^{-7}$ in the big data problem, while these probabilities range between 1.1157×10^{-11} and 2.8407×10^{-1} in the benchmark problem.

In figure 5 we present plots of these probabilities for the two applications considered in sections 5.1 and 5.2. Moreover, some descriptive statistics are shown in table 1.

6. Conclusions

We investigate stochastic LWK 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 a stochastic version of the PLKW method in [23] (see also [24]).

We prove monotonicity of the proposed sPLWK method (proposition 3.3). Moreover, we provide estimates to the *average gain* (lemmas 3.2 and 4.1) as well as a lower bound to the step-sizes λ_{I_k} proposed in (6b) (remark 2.5). A convergence proof in the case of exact data is given (theorem 3.6). In the noisy data case, stability and semiconvergence results are established (theorems 4.3 and 4.4).

An algorithmic implementation of the sPLWK method is presented. The resulting iteration is tested and compared with two well-known Kaczmarz type methods, namely LWK and PLWK.

Two applications are considered: (i) a well-known benchmark problem modeled by a large scale Hilbert type matrix with 10^8 lines; (ii) a big data linear regression problem using real data from a gas sensor array under dynamic gas mixtures [12]. The obtained results validate the efficiency of our method.

It is worth noticing that, in our numerical experiments, the choice of the probabilities $(p_i)_{i=0}^{N-1}$ may strongly influence the performance of the algorithm. Indeed, choosing probabilities proportional to the square of the norm of the rows allows for faster convergence when these norms vary over a wider range. On the other hand, when the range is small, this choice of probabilities are very similar to the uniform distribution and, numerically, we have not observed a improvement over the deterministic version of the algorithm.

Acknowledgments

AL acknowledges support from the National Council for Scientific and Technological Development (CNPq, Grant 311087/2017-5), and from the AvH Foundation.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

A Leitão https://orcid.org/0000-0001-6785-8835

References

[1] Bakushinsky A B and Kokurin M Y 2004 *Iterative Methods for Approximate Solution of Inverse Problems* (*Mathematics and its Applications* vol 577) (Berlin: Springer)

[2] Baumeister J, Kaltenbacher B, Kaltenbacher B and Leitão A 2010 On Levenberg–Marquardt–Kaczmarz iterative methods for solving systems of nonlinear ill-posed equations *Inverse Problem Imaging* **4** 335–50

[3] Bauschke H H and Borwein J M 1997 Legendre functions and the method of random Bregman projections *J. Convex Anal.* **4** 27–67

[4] Browder F E and Petryshyn W V 1967 Construction of fixed points of nonlinear mappings in Hilbert space *J. Math. Anal. Appl.* **20** 197–228

[5] Burger M and Kaltenbacher B 2006 Regularizing Newton–Kaczmarz methods for nonlinear ill-posed problems *SIAM J. Numer. Anal.* **44** 153–82

[6] Byrne C L 2015 *Signal Processing: A mathematical approach* (*Monographs and Research Notes in Mathematics*) 2nd edn (Boca Raton, FL: CRC Press)

[7] Cullen M, Freitag M A, Kindermann S and Scheichl R (ed) 2013 *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences* (*Radon Series on Computational And Applied Mathematics* vol 13) (Berlin: de Gruyter & Co)

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

[9] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)

[10] Filippozzi R, Hafemann E, Rabelo J, Margotti F and Leitão A 2021 A range-relaxed criteria for choosing the Lagrange multipliers in the Levenberg Marquardt Kaczmarz method for solving systems of nonlinear ill-posed equations: application to EIT-CEM with real data http://mtm.ufsc.br/~aleitao/index-ProdJ.html (submitted)

[11] Filippozzi R, Rabelo J C, Boiger R and Leitão A 2021 A range-relaxed criteria for choosing the Lagrange multipliers in the iterated Tikhonov Kaczmarz method for solving systems of linear ill-posed equations *Inverse Problems* **37** 045005

[12] Fonollosa J, Sheik S, Huerta R and Marco S 2015 Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring *Sensors Actuators* B **215** 618–29

[13] Groetsch C W 2007 *Stable Approximate Evaluation of Unbounded Operators* (*Lecture Notes in Mathematics* vol 1894) (Berlin: Springer)

[14] Haltmeier M, Kowar R, Kowar R, Leitão A and Scherzer O 2007 Kaczmarz methods for regularizing nonlinear ill-posed equations: II. Applications *Inverse Problem Imaging* **1** 507–23

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

[16] Haltmeier M, Leitão A, Leitão A and Scherzer O 2007 Kaczmarz methods for regularizing nonlinear ill-posed equations: I. Convergence analysis *Inverse Problem Imaging* **1** 289–98

[17] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of the Landweber iteration for nonlinear ill-posed problems *Numer. Math.* **72** 21–37

[18] Herman G T 1975 A relaxation method for reconstructing objects from noisy x-rays *Math. Program.* **8** 1–19

[19] Herman G T 1980 The fundamentals of computerized tomography *Image Reconstruction from Projections* (New York: Academic) Computer Science and Applied Mathematics

[20] Kaczmarz S 1937 Angenäherte auflösung von systemen linearer gleichungen *Bull. Int. Acad. Polonaise des Sci. Lett.* A **35** 355–7

[21] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (*Radon Series on Computational and Applied Mathematics* vol 6) (Berlin: de Gruyter & Co)

[22] Landweber L 1951 An iteration formula for Fredholm integral equations of the first kind *Am. J. Math.* **73** 615–24

[23] Leitão A and Svaiter B F 2016 On projective Landweber–Kaczmarz methods for solving systems of nonlinear ill-posed equations *Inverse Problems* **32** 025004

[24] Leitão A and Svaiter B F 2018 On a family of gradient-type projection methods for nonlinear ill-posed problems *Numer. Funct. Anal. Optim.* **39** 1153–80

[25] Margotti F, Rieder A and LeitÃo A 2014 A Kaczmarz version of the reginn-Landweber iteration for ill-posed problems in Banach spaces *SIAM J. Numer. Anal.* **52** 1439–65

[26] Morozov V A 1993 *Regularization Methods for Ill-Posed Problems* (Boca Raton, FL: CRC Press)

[27] Natterer F 1977 Regularisierung schlecht gestellter Probleme durch Projektionsverfahren *Numer. Math.* **28** 329–41

[28] Natterer F 1986 *The Mathematics of Computerized Tomography* ed B G Teubner (New York: Wiley)

[29] Natterer F 1997 Algorithms in tomography *The State of the Art in Numerical Analysis* vol 63 (New York: Oxford University Press) pp 503–23

[30] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (Philadelphia, PA: SIAM)

[31] Sabbagh H A, Murphy R K, Sabbagh E H, Aldrin J C and Knopp J S 2013 A modern paradigm for eddy-current nondestructive evaluation *Computational Electromagnetics and Model-Based inversion* (New York: Springer)

[32] Scherzer O 1993 Convergence rates of iterated Tikhonov regularized solutions of nonlinear III? posed problems *Numer. Math.* **66** 259–79

[33] Seidman T I and Vogel C R 1989 Well posedness and convergence of some regularisation methods for non-linear ill posed problems *Inverse Problems* **5** 227–38

[34] Strohmer T and Vershynin R 2009 A randomized Kaczmarz algorithm with exponential convergence *J. Fourier Anal. Appl.* **15** 262–78

[35] Tikhonov A N 1963 Regularization of incorrectly posed problems *Soviet Math. Dokl.* **4** 1624–7

[36] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-Posed Problems* ed F John (New York: Wiley)

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

[38] Wang H, Yang M and Stufken J 2018 Information-based optimal subdata selection for big data linear regression *J. Am. Stat. Assoc.* **114** 393

[39] Bottou L, Curtis F E and Nocedal J 2018 Optimization methods for large-scale machine learning *SIAM Rev.* **60** 223–311

[40] Geiersbach C and Pflug G C 2019 Projected stochastic gradients for convex constrained problems in Hilbert spaces *SIAM J. Optim.* **29** 2079–99

[41] Jin B, Zhou Z and Zou J 2021 On the convergence of stochastic gradient descent for nonlinear ill-posed problems *SIAM J. Optim.* **30** 1421–50

[42] Needell D, Srebro N and Ward R 2016 Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm *Math. Program.* **155** 549–73

[43] Xiao L and Zhang T 2014 A proximal stochastic gradient method with progressive variance reduction *SIAM J. Optim.* **24** 2057–75