IOP Publishing Inverse Problems

Inverse Problems 37 (2021) 045005 (20pp) https://doi.org/10.1088/1361-6420/abc233

A range-relaxed criteria for choosing the Lagrange multipliers in the iterated Tikhonov Kaczmarz method for solving systems of linear ill-posed equations

R Filippozzi[1], J C Rabelo[1,2], R Boiger[3] and A Leitão[1,*]

[1] Department of Mathematics, Federal University of St. Catarina, PO Box 476, 88040-900 Florianópolis SC, Brazil
[2] On leave from Department of Mathematics, Federal University of Piaui, 64049-550, Teresina PI, Brazil
[3] Materials Center Leoben Forschung Gmbh, Roseggerstr. 12, 8700 Leoben, Austria

E-mail: rafaela.filippozzi@gmail.com, joelrabelo@ufpi.edu.br, boiger.romana@gmail.com and acgleitao@gmail.com

Received 8 July 2020, revised 10 September 2020
Accepted for publication 16 October 2020
Published 26 February 2021

CrossMark

Abstract

In this article we propose and analyze a nonstationary *iterated Tikhonov Kaczmarz* (iTK) type method for obtaining stable approximate solutions to systems of ill-posed equations modeled by linear operators acting between Hilbert spaces. We generalize for the iTK iteration the criteria proposed in [5] for the *iterated Tikhonov* method. The goal is to devise an efficient strategy for choosing the Lagrange multipliers in this method. Convergence analysis for the resulting iTK method is provided, including convergence for exact data, stability and semi-convergence. Numerical experiments are presented for two distinct applications, namely: an image deblurring problem and a 2D elliptic parameter identification problem (the inverse potential problem). The obtained numerical results validate the efficiency of the proposed method.

Keywords: ill-posed problems, systems of linear equations, iterated Tikhonov method, Kaczmarz method

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

1. Introduction

In this article we propose a nonstationary *iterated Tikhonov Kaczmarz* (iTK) type method for obtaining regularized approximations of systems of linear ill-posed operator equations. This

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

is a Kaczmarz type method [25], where each step is defined as in the iterated Tikhonov (iT) method [6, section 1.2] with the Lagrange multiplier being chosen as to guarantee the residual of the next iterate to be in a *range* [5].

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 and Y are Hilbert spaces, and $N \geqslant 1$. In practical situations, one does not know the data exactly. Instead, only approximate measured data $y_i^\delta \in Y$ satisfying

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

are available, where $\delta_i > 0$ are the (known) noise levels. The available data y_i^δ are obtained by indirect measurements of the parameter x. This process being described by the system of ill-posed operator equations

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

where $A_i : X \to Y$ are bounded linear operators, whose inverses $A_i^{-1} : R(A_i) \to X$ either do not exist, or are not continuous. Consequently, approximate solutions are extremely sensitive to noise in the data.

Linear ill-posed problems are commonly found in applications, ranging from image analysis to parameter identification in mathematical models. There is a vast literature on iterative methods for the stable solution of (2). We refer the reader to the text books [1, 2, 13, 16, 23, 26, 27, 30, 31, 33] and the references therein.

1.1. Iterated Tikhonov type methods

Standard iT type methods for solving the ill-posed problem (1) and (2) are defined, after rewriting (2) as a single equation $\mathbf{A}x = \mathbf{y}$, where $\mathbf{A} = (A_0, \ldots, A_{N-1}) : X \to Y^N$ and $\mathbf{y}^\delta = (y_0^\delta, \ldots, y_{N-1}^\delta)$, by the iteration formula

$$x_{k+1}^\delta = \arg\min_{x \in X} \left\{ \lambda_k \|\mathbf{A}x - \mathbf{y}^\delta\|^2 + \|x - x_k^\delta\|^2 \right\} \tag{3}$$

or, equivalently, by

$$\begin{aligned} x_{k+1}^\delta &= x_k^\delta - \lambda_k \left(I + \lambda_k \mathbf{A}^*\mathbf{A}\right)^{-1} \mathbf{A}^* \left(\mathbf{A}x_k^\delta - \mathbf{y}^\delta\right) \\ &= \left(\lambda_k^{-1} I + \mathbf{A}^*\mathbf{A}\right)^{-1} \left[\lambda_k^{-1} x_k^\delta + \mathbf{A}^*\mathbf{y}^\delta\right], \end{aligned} \tag{4}$$

where $\mathbf{A}^* : Y^N \to X$ is the adjoint operator of $\mathbf{A}$. The parameter $\lambda_k > 0$ can be viewed as the Lagrange multiplier of the problem of projecting x_k^δ onto a levelset of $\|\mathbf{A}x - \mathbf{y}^\delta\|^2$ (see [5, section 1, pg.4] or [13, section 5.1, pg 122]). If the sequence $\{\lambda_k = \lambda\}$ is constant, iteration (4) is called *stationary* iT [16, 28, 30], otherwise it is denominated *nonstationary* iT [6, 14, 20].

In the nonstationary iT methods, each λ_k is chosen either *a priori* (e.g., the geometrical choice $\lambda_k = q^k$, $q > 1$) or *a posteriori* [5, 11]. In this article we focus on the *a posteriori* strategy investigated in [5], where the authors propose a choice for the Lagrange multipliers, which requires the residual at the next iterate to assume a prescribed value dependent on the current residual and also on the noise level. More precisely, λ_k is chosen so that the next iterate has a prescribed residual satisfying $\delta \leqslant \|\mathbf{A}x_{k+1}^\delta - \mathbf{y}^\delta\| \leqslant \Phi(\|\mathbf{A}x_k^\delta - \mathbf{y}^\delta\|, \delta)$, where Φ represents a convex combination of $\|\mathbf{A}x_k^\delta - \mathbf{y}^\delta\|$ and δ.

The iT type methods may become inefficient if N is large or the evaluation of the step in (4) is expensive. In such cases, Kaczmarz type methods which cyclically consider each equation in (2) separately, are reported to be faster [32] and are often the method of choice in practice. On the other hand, only few theoretical results about regularizing properties of iT-Kaczmarz methods are available, so far (see, e.g., [9]).

1.2. Iterated Tikhonov Kaczmarz type methods

The method proposed and analyzed in this manuscript for solving the ill-posed problem (1) and (2) is a Kaczmarz type method, where each step is defined as in the iT method (4) and the choice of Lagrange multipliers proposed in [5] is adopted. This iterative method is defined by

$$x_{k+1}^{\delta} \;=\; x_k^{\delta} + h_k, \tag{5}$$

where

$$h_k \;=\; \begin{cases} \lambda_k (I + \lambda_k A_{[k]}^* A_{[k]})^{-1} A_{[k]}^* \left(y_{[k]}^{\delta} - A_{[k]} x_k^{\delta} \right), & \text{if } \|A_{[k]} x_k^{\delta} - y_{[k]}^{\delta}\| > \tau \delta_{[k]} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

and

$$\lambda_k \;=\; \begin{cases} \text{chosen as in algorithm 1}, & \text{if } \|A_{[k]} x_k^{\delta} - y_{[k]}^{\delta}\| > \tau \delta_{[k]} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Here $[k] = (k \bmod N) \in \{0, 1, \ldots, N-1\}$, $x_0^{\delta} = x_0 \in X$ is an initial guess and $\tau > 1$ is a fixed constant (see next section).

The $h_k \in X$ in (6) is inspired in the iterative step proposed in [5] for the $[k]$-th equation $A_{[k]} x = y_{[k]}$ of system (2), with data $y_{[k]}^{\delta}$ given as in (1). Notice that, if $\|A_{[k]} x_k^{\delta} - y_{[k]}^{\delta}\| \leqslant \tau \delta_{[k]}$, for some k, then the computation of (λ_k, h_k) is avoided and we set $\lambda_k = 0$, $h_k = 0$ and $x_{k+1}^{\delta} = x_k^{\delta}$.

Following [5] we refer to this method as *range-relaxed iterated Tikhonov Kaczmarz* (rriTK) method. Essentially, it consists in incorporating the Kaczmarz strategy to the range-relaxed iterated Tikhonov (rriT) method in [5]. This procedure is analog to the one introduced in [10, 17–19], and [7] regarding the Landweber Kaczmarz (LWK), the steepest descent Kaczmarz, the expectation maximization Kaczmarz, and the iteratively regularized Gauss–Newton–Kaczmarz (irGNK) iterations respectively. It is worth mentioning that iteration (5)–(7) was considered in [9] for the constant choice $\lambda_k = \lambda$.

In Kaczmarz type algorithms, a group of N subsequent steps (starting at some multiple of N) is called a cycle. The iteration (5)–(7) should be terminated when, for the first time, all x_k^{δ} are equal within a cycle. That is, we stop the iteration at step $k_* = k_* \left(\{\delta_i\}_i, \{y_i^{\delta}\}_i \right)$ s.t.

$$k_* \;:=\; \min \left\{ lN : l \in \mathbb{N} \quad \text{and} \quad x_{lN}^{\delta} = x_{lN+1}^{\delta} = \cdots = x_{lN+N}^{\delta} \right\}. \tag{8}$$

In other words, $k_* \in \mathbb{N}$ is the smallest multiple of N such that $x_{k_*}^{\delta} = x_{k_*+1}^{\delta} = \cdots = x_{k_*+N}^{\delta}$ or, equivalently, such that $\lambda_{k_*} = \lambda_{k_*+1} = \cdots = \lambda_{k_*+N} = 0$.

1.3. Outline of the manuscript

The article is organized as follows: in section 2 we introduce the rriTK method, proposed and analyzed in this manuscript. A detailed formulation of this method is given. It is proven that the method is well defined and some preliminary results are obtained, including an estimate for the '*gain*' (proposition 2.4), as well as an estimate for the Lagrange multipliers λ_k (corollary 2.6). In section 3 a convergence result for the exact data case is presented (theorem 3.2). Stability and semi-convergence results are established in section 4 (theorems 4.3 and 4.4 respectively). Section 5 is devoted to numerical experiments. Two distinct applications are considered: an image deblurring problem and a 2D elliptic parameter identification problem (the inverse potential problem (IPP)). The performance of the rriTK method is compared with other Kaczmarz type methods, namely the geometric iTK method (giTK) with $\lambda_k = 2^k$, the stationary iTK method (siTK) with $\lambda_k = 2$ and the Landweber-Kaczmarz method (LWK). Section 6 is dedicated to final remarks and conclusions.

2. A range-relaxed iterated Tikhonov Kaczmarz method

In the following we introduce the rriTK method for solving the ill-posed linear system (1) and (2). Subsection 2.1 is devoted to main assumptions needed in the analysis. The new method is presented in subsection 2.2 and a corresponding algorithm is discussed. In subsection 2.3 we derive some basic properties of the proposed method, and prove preliminary results and estimates.

The implementable method proposed here, happens to be a nonstationary iTK type method where, in each iteration, the set of feasible choices for the Lagrange multipliers is an interval, instead of a single real number. For this reason, this method is called a (nonstationary) range-relaxed iT method.

2.1. Main assumptions

For the remaining of this article we suppose that the following assumptions hold true:

(A1) There exists $x^\star \in X$ such that $A_i x^\star = y_i$, where $y_i \in R(A_i)$, $i = 0, \ldots, N-1$, are the exact data.

(A2) The operators $A_i : X \to Y$ are linear, bounded and ill-posed, 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.

From (A2) it follows the existence of $C > 0$ with $\max_i \|A_i\| \leqslant C$.

2.2. Description of the method

As already discussed in the introduction, the iterative step of the rriTK method is analog to the one proposed in [5]. This step is discussed in the following.

Given $k \in \mathbb{N}$, set $i = [k]$ and define for $\mu > 0$ the levelsets $\Omega^i_\mu := \{x \in X;\ \|A_i x - y_i^\delta\| \leqslant \mu\}$ of the residual w.r.t. the ith-equation of system (2). If the iterate x_k^δ does not belong to $\Omega^i_{\delta_i}$, the next iterate x_{k+1}^δ is computed by solving the *range-relaxed projection problem*

$$\begin{cases} \min\limits_x \ \|x - x_k^\delta\|^2 \\ \text{s.t. } \|A_i x - y_i^\delta\|^2 \leqslant \mu^2, \quad \bar{\Phi}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i) \leqslant \mu \leqslant \bar{\bar{\Phi}}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i) \end{cases} \tag{9}$$

for $(x, \mu) \in X \times \mathbb{R}$, where $\bar{\Phi}(u, v) = \bar{p}u + (1 - \bar{p})v$ and $\bar{\bar{\Phi}}(u, v) = \bar{\bar{p}}u + (1 - \bar{\bar{p}})v$, $\forall\ u, v \in \mathbb{R}$, with $0 < \bar{p} < \bar{\bar{p}} < 1$. Thus the interval $[\bar{\Phi}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i), \bar{\bar{\Phi}}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i)]$ is non-degenerate. If (x', μ') is a solution of (9), we define $x_{k+1}^\delta = x'$ and $\|A_i x_{k+1}^\delta - y_i^\delta\| = \mu'$ (see lemma 2.1). As observed in [5], x_{k+1}^δ is generated from x_k^δ by projecting it onto anyone of the range of convex sets $(\Omega^i_\mu)_{\bar{\Phi} \leqslant \mu \leqslant \bar{\bar{\Phi}}}$.

Algorithm 1. The range-relaxed iterated Tikhonov Kaczmarz (rriTK) method.

[1] choose an initial guess $x_0 \in X$ and $\lambda_{\max} > 0$; set $k := 0$;
[2] choose $\tau > 1$, and $0 < \bar{p} < \bar{\bar{p}} < 1$;
[3] **repeat**
[3.1] $i = [k]$;
[3.2] **if** $\left[\ \|A_i\, x_k^\delta - y_i^\delta\| \ > \ \tau\delta_i \ \right]$ **then**
compute $(\lambda_k, h_k) \in \mathbb{R} \times X$ such that

$$\begin{cases} h_k, = \ -\lambda_k \left(I + \lambda_k A_i^* A_i\right)^{-1} A_i^* \left(A_i\, x_k^\delta - y_i^\delta\right) \\ \bar{p}\|A_i x_k^\delta - y_i^\delta\| + (1-\bar{p})\delta_i \ \leqslant \ \|A_i(x_k^\delta + h_k) - y_i^\delta\| \ \leqslant \ \bar{\bar{p}}\|A_i x_k^\delta - y_i^\delta\| + (1-\bar{\bar{p}})\delta_i \end{cases}$$

if$[\lambda_k \ > \ \lambda_{\max}]$**then**

$\lambda_k = \lambda_{\max}$; $h_k \ = \ -\lambda_{\max} \ \left(I + \lambda_{\max}\ A_i^* A_i\right)^{-1} A_i^* \left(A_i\, x_k^\delta - y_i^\delta\right)$

else

$\lambda_k = 0$; $h_k = 0$;
[3.3] $x_{k+1}^\delta \ = \ x_k^\delta + h_k$;
[3.4] $k = k + 1$;
until $[([k] = 0)$ and $(\lambda_{k-1} = \lambda_{k-2} = \cdots = \lambda_{k-N} = 0)]$;
[4] $k_* = k - N$;

Since the solution of (9) is not unique, there are several possible choices for x_{k+1}^δ. The next lemma addresses this problem. For a proof we refer the reader to [5, lemma 2.3].

Lemma 2.1. *Suppose* $\|A_i x_k^\delta - y_i^\delta\| > \delta_i$. *The following assertions are equivalent:*

(*a*) $x' = \Pi_{\Omega_\mu}(x_k^\delta)$ *and* $\bar{\Phi}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i) \leqslant \mu' \leqslant \bar{\bar{\Phi}}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i)$;

(*b*) $(x', \mu') \in X \times \mathbb{R}$ *is a solution of the range-relaxed projection problem* (9);

(*c*) $x' = x_k^\delta - \lambda(I + \lambda A_i^* A_i)^{-1} A_i^* (A_i x_k^\delta - y_i^\delta)$, *for some* $\lambda > 0$,

$$\bar{\Phi}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i) \leqslant \|A_i x' - y_i^\delta\| \leqslant \bar{\bar{\Phi}}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i),$$

and $\mu' = \|A_i x' - y_i^\delta\|$;

(*here* $\Pi_\Omega(x)$ *represents the orthogonal projection of* x *onto the convex set* Ω).

It follows from lemma 2.1 that solving the range-relaxed projection problem in (9) sums up to solving the inequalities $\bar{\Phi}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i) \leqslant \|A_i x' - y_i^\delta\| \leqslant \bar{\bar{\Phi}}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i)$, where $x' = x_k^\delta - \lambda(I + \lambda A_i^* A_i)^{-1} A_i^* (A_i x_k^\delta - y_i^\delta)$, and $\mu' = \|A_i x' - y_i^\delta\|$.

We use this result to propose an implementable version of the rriTK method in algorithm 1.

Remark 2.2 (On the bound of the Lagrange multipliers in algorithm 1).

- In algorithm 1 the Lagrange multipliers λ_k are bounded from above by some $\lambda_{\max} > 0$. This is needed in order to prove convergence for exact data (see theorem 3.2).

This assumption is also used in the proof of the stability theorem 4.3 (see step 1 of the proof). However, an alternative proof of theorem 4.3 can be given without this assumption.

- It is worth noticing that, in the noisy data case the assumption $\lambda_k \leqslant \lambda_{\max}$, $k = 0, \ldots, k^*$, plays no role, since $\lambda_{\max}$ can be chosen arbitrarily large and algorithm 1 always stops after a finite number of steps (see corollary 2.8).

Consequently, the rriTK method can be implemented without any bound on the λ_k's. [4]

[4] As a matter fact, it is to expect that the multipliers will assume larger values for problems with small levels of noise (see numerical experiments in section 5).

- In the particular case $N = 1$, system (2) reduces to a single linear operator equation, and all results in this article can be proven without the assumption $\lambda_k \leqslant \lambda_{\max}$. In this case, the rriTK investigated in this manuscript happens to be a particular instance of the rrLM method considered in [29] for solving nonlinear operator equations.

2.3. Preliminary results

For simplicity of notation we write $b_k^\delta := y_i^\delta - A_i x_{k+1}^\delta = y_i^\delta - A_i h_k - A_i x_k^\delta$, with $i = [k]$, and $C := \max_j \|A_j\|$. Moreover, for exact data $y = (y_0, \ldots, y_{N-1})$, the iterates in (5) are denoted by x_k, in contrast to x_k^δ in the noisy data case (analog notation for $b_k := y_i - A_i x_{k+1}$).

Our first result concerns basic properties of the iterative step of the rriTK method. The proofs of the assertions are straightforward and will be omitted.

Lemma 2.3. *Assume that* (*A1*) *and* (*A2*) *are satisfied and let* x_k^δ, h_k, λ_k *be defined by* (5)–(7) *respectively. In the noisy data case, if* $\lambda_{\max}$ *is large enough, the assertions*

(*a*) $A_i\, x_{k+1}^\delta - y_i^\delta \; = \; (\lambda_k A_i A_i^* + I)^{-1}(A_i\, x_k^\delta - y_i^\delta)$*;*
(*b*) $h_k \; = \; \lambda_k A_i^* \left(y_i^\delta - A_i\, x_{k+1}^\delta\right)$*;*
(*c*) $\underline{p}\|A_i\, x_k^\delta - y_i^\delta\| \leqslant \|A_i\, x_{k+1}^\delta - y_i^\delta\| \leqslant \|A_i\, x_k^\delta - y_i^\delta\|$*;*

hold true for $0 \leqslant k < k_*$. *In the exact data case assertions a*) *and b*) *hold true. Moreover,*

(*d*) $\underline{p}\|A_i\, x_k - y_i\| \leqslant \|A_i\, x_{k+1} - y_i\| \leqslant \overline{p}\|A_i\, x_k - y_i\|$, *whenever* $\lambda_k < \lambda_{\max}$*;*
(*e*) $(1 + \lambda_{\max}\, C^2)^{-1}\|A_i\, x_k - y_i\| \leqslant \|A_i\, x_{k+1} - y_i\| \leqslant \|A_i\, x_k - y_i\|$, *whenever* $\lambda_k = \lambda_{\max}$.

In what follows we estimate the 'gain' $\|x_{k+1}^\delta - x^\star\|^2 - \|x_k^\delta - x^\star\|^2$. This is a central result for the analysis derived in this manuscript.

Proposition 2.4. *Assume that* (*A1*) *and* (*A2*) *are satisfied and let* x_k^δ, h_k, λ_k *be defined by* (5)–(7) *respectively. Then*

$$\|x_{k+1}^\delta - x^\star\|^2 - \|x_k^\delta - x^\star\|^2 \;\leqslant\; 2\overline{p}\,\lambda_k\,\|b_k^\delta\|(\delta_i - \|A_i x_k^\delta - y_i^\delta\|) - \|x_{k+1}^\delta - x_k^\delta\|^2, \quad (10)$$

for $k = 0, \ldots, k_* - 1$. *In particular, in the exact data case* ($y_i^\delta = y_i$) *we have*

$$\|x_{k+1} - x^\star\|^2 - \|x_k - x^\star\|^2 \leqslant -2\gamma^2\,\lambda_k\|A_i\, x_k - y_i\|^2 - \|x_{k+1} - x_k\|^2, \qquad (11)$$

for $k = 0, \ldots,$ *with* $\gamma := \;\min\{\underline{p}, (1 + \lambda_{\max}\; C^2)^{-1}\}$.

Proof. Let $i = [k]$. If $\|A_i\, x_k^\delta - y_i^\delta\| \leqslant \tau\delta_i$, then $\lambda_k = 0$ and $x_{k+1}^\delta = x_k^\delta$. Thus, (10) is trivial. Otherwise, it follows from lemma 2.3(b) that

$$\begin{aligned}
&\|x_{k+1}^\delta - x^\star\|^2 - \|x_k^\delta - x^\star\|^2 \\
&\quad= \;2\,\langle x_{k+1}^\delta - x_k^\delta,\; x_{k+1}^\delta - x^*\rangle - \|x_{k+1}^\delta - x_k^\delta\|^2 \\
&\quad= \;2\lambda_k\,\langle y_i^\delta - A_i x_{k+1}^\delta,\; A_i(x_{k+1}^\delta - x^\star)\rangle - \|x_{k+1}^\delta - x_k^\delta\|^2 \\
&\quad= \;2\lambda_k\,\langle y_i^\delta - A_i x_{k+1}^\delta,\; A_i x_{k+1}^\delta - y_i^\delta + y_i^\delta - A_i x^\star\rangle - \|x_{k+1}^\delta - x_k^\delta\|^2 \\
&\quad\leqslant \;2\lambda_k\,\left[-\|b_k^\delta\|^2 + \|b_k^\delta\|\,\delta_i\right] - \|x_{k+1}^\delta - x_k^\delta\|^2. \qquad (12)
\end{aligned}$$

It follows from step [3.2] of algorithm 1 that $\overline{\Phi}(\|A_i x_k^\delta - y_i^\delta\|, \delta_i) = \overline{p}\|A_i x_k^\delta - y_i^\delta\| + (1 - \overline{p})\delta_i \leqslant \|b_k^\delta\|$. Consequently $-\|b_k^\delta\|^2 + \|b_k^\delta\|\delta_i = \|b_k^\delta\|(\delta_i - \|b_k^\delta\|) \leqslant \overline{p}\|b_k^\delta\|(\delta_i - \|A_i x_k^\delta - y_i^\delta\|)$, and (10) follows. In the exact data case, it follows from lemma 2.3(d) and (e) that $-\|b_k\|^2 \leqslant -\min\{\bar{p}^2, (1+\lambda_{\max}\, C^2)^{-2}\}\|A_i x_k - y_i\|^2$; consequently (12) implies (11). $\square$

Proposition 2.4 has several relevant consequences, namely: monotonicity of the rriTK method (corollary 2.5); a uniform estimate for the Lagrange multipliers (corollary 2.6); summability of important series (corollary 2.7); finiteness of the stoping index k_* (corollary 2.8).

Corollary 2.5. *Assume that (A1) and (A2) are satisfied and let x_k^δ, h_k, λ_k be defined by (5)–(7) respectively. Then*

$$\|x_{k+1}^\delta - x^\star\|^2 \ \leqslant\ \|x_k^\delta - x^\star\|^2, \quad k = 0, \ldots, k_* - 1. \tag{13}$$

Additionally, in the exact data case we have $\|x_{k+1} - x^\star\|^2 \leqslant \|x_k - x^\star\|^2$, for $k = 0, 1, \ldots$.

Corollary 2.6. *Assume that (A1) and (A2) are satisfied and let x_k^δ, h_k, λ_k be defined by (5)–(7) respectively. Moreover, let b_k^δ be defined as above. Then*

$$\lambda_k \ \geqslant\ \frac{\left(\|A_i x_k^\delta - y_i^\delta\| - \|b_k^\delta\|\right)\ \|A_i x_k^\delta - y_i^\delta\|}{\|A_i^*(A_i x_k^\delta - y_i^\delta)\|^2}, \quad k = 0, \ldots, k_* - 1. \tag{14}$$

Moreover, if $\|A_i\, x_k^\delta - y_i^\delta\| > \tau\delta_i$ (for some $0 \leqslant k \leqslant k_* - 1$) then $\lambda_k > C^{-2}(1-\bar{\bar{p}})(1-1/\tau)$, with $C > 0$ defined as above.

Additionally, in the exact data case it holds $\lambda_k \geqslant\ \min\{\lambda_{\max}, C^{-2}(1-\bar{\bar{p}})\}$.

Proof. Let $0 \leqslant k < k_*$. If $\|A_i x_k^\delta - y_i^\delta\| \leqslant \tau\delta_i$ then $\lambda_k = 0$ and $x_{k+1}^\delta = x_k^\delta$. Thus, (14) is trivial. On the other hand, if $\|A_i x_k^\delta - y_i^\delta\| > \tau\delta_i$, the proof of (14) follows the lines of [5, corollary 2.5].

To prove the second assertion, notice that $\|b_k^\delta\| \leqslant \bar{\bar{p}}\|A_i x_k^\delta - y_i^\delta\| + (1-\bar{\bar{p}})\delta$. Consequently, $\|A_i x_k^\delta - y_i^\delta\| - \|b_k^\delta\| \geqslant (1-\bar{\bar{p}})(\|A_i x_k^\delta - y_i^\delta\| - \delta_i)$. Thus, it follows from (14) that

$$\begin{aligned}\lambda_k &\geqslant C^{-2}\frac{\|A_i x_k^\delta - y_i^\delta\| - \|b_k^\delta\|}{\|A_i x_k^\delta - y_i^\delta\|} \\ &\geqslant C^{-2}(1-\bar{\bar{p}})\left(1 - \frac{\delta_i}{\|A_i x_k - y_i\|}\right) > C^{-2}(1-\bar{\bar{p}})(1-1/\tau).\end{aligned}$$

In the exact data case, given $k \in \mathbb{N}$, either $\lambda_k < \lambda_{\max}$ and $\|b_k\| \leqslant \bar{\bar{p}}\|A_i x_k - y_i\|$ (see lemma 2.3(d)) or $\lambda_k = \lambda_{\max}$ (see lemma 2.3(e)). In the first case, it follows from (14) that

$$\lambda_k \geqslant C^{-2}\left(1 - \frac{\|b_k\|}{\|A_i x_k - y_i\|}\right) \geqslant C^{-2}(1-\bar{\bar{p}}).$$

Consequently, $\lambda_k \geqslant\ \min\{\lambda_{\max}, C^{-2}(1-\bar{\bar{p}})\}$. $\square$

Corollary 2.7. *Assume that (A1) and (A2) are satisfied and let x_k, h_k, λ_k be defined by (5)–(7) in the exact data case (i.e. $y_i^\delta = y_i$, $i = 0, \ldots, N-1$). Then the series*

$$\sum_{k=0}^{\infty}\|x_{k+1} - x_k\|^2, \qquad \sum_{k=0}^{\infty}\lambda_k\,\|A_{[k]}x_k - y_{[k]}\|^2,$$

$$\sum_{k=0}^{\infty}\lambda_k\,\|b_k\|^2 \quad \textit{and} \quad \sum_{k=0}^{\infty}\|A_{[k]}x_k - y_{[k]}\|^2$$

are all summable.

Proof. The first two assertions follow from (11), using a telescopic series argument. The next assertion follow from a comparison test and lemma 2.3(d) and (e). The last assertion follows from the second one and corollary 2.6. □

Corollary 2.8. *Assume that* (*A1*) *and* (*A2*) *are satisfied and let* x_k^δ, h_k, λ_k *be defined by* (5)–(7). *Then the stopping index* k_* *defined in* (8) *is finite and*

$$k_* \leqslant N\|x_0 - x^\star\|^2 \left[2\bar{p}^2 (1 - \bar{\bar{p}})\, C^{-2}\, \delta_{\min}^2 (\tau - 1)^2\right]^{-1}. \tag{15}$$

Proof. Assume by contradiction that k_* is not finite, i.e., in each cycle $\{lN, \ldots, lN + N - 1\}$, $l \in \mathbb{N}$, of the rriTK method, there exists at least one index $j(l) \in \{0, \ldots, N - 1\}$ such that $\|A_{j(l)}\, x_{lN+j(l)} - y_{j(l)}^\delta\| \geqslant \tau\delta_{j(l)}$.

From proposition 2.4 it follows that (10) holds for $k \in \mathbb{N}$. Summing over k and using the fact that either $\|A_{[k]}x_k^\delta - y_{[k]}^\delta\| \geqslant \tau\delta_{[k]}$ or $\lambda_k = 0$, we obtain (with the notation $i = [k]$)

$$\begin{aligned}
\|x_0 - x^*\|^2 &\geqslant 2\bar{p}\sum_{k=0}^{lN} \lambda_k \|b_k^\delta\| \left(\|A_i x_k^\delta - y_i^\delta\| - \delta_i\right) \\
&\geqslant 2\bar{p}^2 \sum_{k=0}^{lN} \lambda_k \|A_i\, x_k^\delta - y_i^\delta\| \left(\|A_i\, x_k^\delta - y_i^\delta\| - \delta_i\right) \\
&\geqslant 2\bar{p}^2 \sum_{s=0}^{l} \lambda_{sN+j(s)} \|A_{j(s)}\, x_{sN+j(s)}^\delta - y_{j(s)}^\delta\| \\
&\quad \times \left(\|A_{j(s)}\, x_{sN+j(s)}^\delta - y_{j(s)}^\delta\| - \delta_{j(s)}\right) \\
&\geqslant 2\bar{p}^2 \sum_{s=0}^{l} \lambda_{sN+j(s)} \tau \delta_{j(s)}^2 (\tau - 1) \\
&\geqslant l\, 2\bar{p}^2 (1 - \bar{\bar{p}}) C^{-2}\, \delta_{\min}^2 (\tau - 1)^2,
\end{aligned} \tag{16}$$

(the last inequality follows from corollary 2.6). Since the right-hand side of (16) becomes unbounded as $l \to \infty$ a contradiction is established, and the finiteness of k_* follows. Estimate (15) follows now substituting $l = k_*/N$ in (16). □

3. A convergence result for exact data

Our main goal in this section is to prove convergence of the rriTK method in the case $\delta_i = 0$, $i = 0, \ldots, N - 1$. Notice that, in this exact data case, $\lambda_k = 0$ and $h_k = x_{k+1} - x_k = 0$ if and only if $\|A_i x_k - y_i\| = 0$ (see step [3.2] of algorithm 1).

Remark 3.1. It is worth noticing that there exists an x_0-minimal norm solution of (2), a solution $x^\dagger$ of (2) such that $\|x^\dagger - x_0\| = \inf\{\|x - x_0\|; \mathbf{A}x = \mathbf{y}\}$. This assertion is a direct consequence of [13]. Moreover, $x^\dagger$ is the only solution of (2) with this property.

Theorem 3.2 (Convergence for exact data). *Assume that* (*A1*) *and* (*A2*) *are satisfied and let* x_k, h_k, λ_k *be defined by* (5)–(7) *in the exact data case* (*i.e.* $y_i^\delta = y_i$, $i = 0, \ldots, N - 1$). *Then* $x_k \to x^\dagger$, *as* $k \to \infty$.

Proof. We define $e_k := x^\star - x_k$. From corollary 2.5 follows that $\|e_k\|$ is monotone non-increasing. Thus, $\|e_k\|$ converges to some $\epsilon \geqslant 0$. In what follows we show that e_k is in fact a Cauchy sequence.

In order to prove that e_k is indeed a Cauchy sequence, it suffices to prove $|\langle e_n - e_k, e_n\rangle| \to 0$, $|\langle e_n - e_l, e_n\rangle| \to 0$ as $k, l \to \infty$ with $k \leqslant l$ for some $k \leqslant n \leqslant l$ [21, theorem 2.3]. Let $k \leqslant l$ be arbitrary and write $k = k_0 N + k_1$, $l = l_0 N + l_1$, with $k_1, l_1 \in \{0, \ldots, N-1\}$. Now let $n_0 \in \{k_0, \ldots, l_0\}$ be such that for all $i_0 \in \{k_0, \ldots, l_0\}$

$$\sum_{s=0}^{N-1} \lambda_{n_0 N+s} \|A_s\, x_{n_0 N+s} - y_s\| \ \leqslant\ \sum_{s=0}^{N-1} \lambda_{i_0 N+s} \|A_s\, x_{i_0 N+s} - y_s\|, \tag{17}$$

and set $n := n_0 N + N - 1$ (if $n_0 = l_0$, we set $n := n_0 N + l_1$; so that $k \leqslant n \leqslant l$). Therefore

$$\begin{aligned} |\langle e_n - e_k, e_n\rangle| &= \left|\sum_{i=k}^{n-1} \langle (x_{i+1} - x_i), (x_n - x^\star)\rangle\right| \\ &= \left|\sum_{i=k}^{n-1} \lambda_i \langle A_{[i]}\, x_{i+1} - y_{[i]}, A_{[i]}\, x_n - A_{[i]}\, x^\star\rangle\right| \\ &\leqslant \sum_{i_0=k_0}^{n_0} \sum_{i_1=0}^{N-1} \lambda_i \|A_{i_1}\, x_{i+1} - y_{i_1}\|\, \|A_{i_1}\, x_n - y_{i_1}\| \\ &\leqslant \sum_{i_0=k_0}^{n_0} \sum_{i_1=0}^{N-1} \lambda_i \|b_i\|\, \|A_{i_1}\, x_n - y_{i_1}\| \end{aligned} \tag{18}$$

(we use the notation $i = i_0 N + i_1$). The last term on the right-hand side of (18) can be estimated by

$$\begin{aligned} \|A_{i_1}\, x_n - y_{i_1}\| &= \|A_{i_1}\, x_{n_0 N+N-1} - y_{i_1}\| \\ &\leqslant \|A_{i_1}\, x_{n_0 N+i_1+1} - y_{i_1}\| + \sum_{s=i_1+1}^{N-2} \|A_{i_1}\, x_{n_0 N+s+1} - A_{i_1}\, x_{n_0 N+s}\| \\ &\leqslant \|A_{i_1}\, x_{n_0 N+i_1+1} - y_{i_1}\| + \sum_{s=i_1+1}^{N-2} C\|x_{n_0 N+s+1} - x_{n_0 N+s}\| \\ &\leqslant \|A_{i_1}\, x_{n_0 N+i_1+1} - y_{i_1}\| + \sum_{s=i_1+1}^{N-2} C\lambda_{n_0 N+s}\|A_s^*(y_s - A_s x_{n_0 N+s})\| \\ &\leqslant \|A_{i_1}\, x_{n_0 N+i_1+1} - y_{i_1}\| + \sum_{s=0}^{N-1} C^2 \lambda_{n_0 N+s}\|A_s x_{n_0 N+s} - y_s\| \\ &\leqslant \sum_{s=0}^{N-1} (1 + C^2 \lambda_{n_0 N+s})\|A_s x_{n_0 N+s} - y_s\| \\ &\leqslant \left(\frac{1}{\lambda_{\min}} + C^2\right) \sum_{s=0}^{N-1} \lambda_{n_0 N+s}\|A_s x_{n_0 N+s} - y_s\|. \end{aligned}$$

(with $\lambda_{\min} = \min\{\lambda_{\max}, C^{-2}(1-\bar{\bar{p}})\}$, see corollary 2.6). Hence, it follows from (17) that $\|A_{i_1}x_n - y_{i_1}\| \leqslant (\frac{1}{\lambda_{\min}} + C^2)\sum_{s=0}^{N-1}\lambda_{i_0N+s}\|A_s x_{i_0N+s} - y_s\|$, for all $i_0 \in \{k_0, \ldots, l_0\}$. Inserting this last inequality into (18) we obtain

$$\begin{aligned} |\langle e_n - e_k, e_n\rangle| &\leqslant \left(\frac{1}{\lambda_{\min}} + C^2\right)\sum_{i_0=k_0}^{n_0}\left[\sum_{i_1=0}^{N-1}\lambda_i\|b_i\|\right]\left[\sum_{s=0}^{N-1}\lambda_{i_0N+s}\|A_s x_{i_0N+s} - y_s\|\right] \\ &= \left(\frac{1}{\lambda_{\min}} + C^2\right)\sum_{i_0=k_0}^{n_0}\left[\sum_{i_1=0}^{N-1}\lambda_i\|b_i\|\right]\left[\sum_{i_1=0}^{N-1}\lambda_i\|A_{i_1}x_i - y_{i_1}\|\right] \\ &\leqslant \left(\frac{1}{\lambda_{\min}} + C^2\right)\frac{N\lambda_{\max}}{2}\sum_{i_0=k_0}^{n_0}\left[\sum_{i_1=0}^{N-1}\lambda_i\|b_i\|^2 + \sum_{i_1=0}^{N-1}\lambda_i\|A_{i_1}x_i - y_{i_1}\|^2\right]. \end{aligned} \tag{19}$$

Hence by corollary 2.7 the right-hand side of (19) goes to zero as $k, l \to \infty$. Analogously one shows that $|\langle e_n - e_l, e_n\rangle| \to 0$ as $k, l \to \infty$.

Thus, e_k is a Cauchy sequence and x_k converges to some $x^+ \in X$. Since the residuals $\|A_{[k]}x_k - y_{[k]}\|$ converge to zero as $k \to \infty$ (see corollary 2.7), this x^+ is a solution of (2).

It follows from lemma 2.3(b) that $x_{k+1} - x_k \in \mathcal{R}(A_i^*) \subset \mathcal{N}(A_i)^\perp \subset \mathcal{N}(\mathbf{A})^\perp$. An inductive argument shows that $x^+ \in x_0 + \mathcal{N}(\mathbf{A})^\perp$, by remark 3.1$x^\dagger$ is the only solution of (2), and with that follows the result. □

4. Convergence for noisy data

In this section we assume that A1 and A2 hold true and that x_k^δ, h_k defined in (5), τ, $\bar{p}$ and $\bar{\bar{p}}$ defined in algorithm 1. Our main goal in this section is to prove that $x_{k_*(\delta)}^\delta$ converges to a solution of (2) as $\delta \to 0$, where $k_*(\delta) = k_*$ is defined in (8). In addition, we will show first that our method is stable.

Definition 4.1. For $k < k_*$, a vector $z \in X$ is called a **successor** of x_k^δ if there exists a pair $(\lambda_k \geqslant 0, h_k \in X)$ defined as in step [3.2] of algorithm 1, such that $z = x_k^\delta + h_k$.

Definition 4.2. A **noiseless sequence** is a sequence $(x_k)_{k\geqslant 0} \subset X$ generated by algorithm 1 with $\delta_i = 0, i = 0, \ldots, N-1$. Notice that

(a) x_{k+1} is a successor of x_k for all $k \in \mathbb{N}$;
(b) $x_{k+1} = \arg\min_x T_{k,\lambda_k}(x)$, where $T_{k,\lambda_k}(x) := \lambda_k\|A_i x - y_i\|^2 + \|x - x_k\|^2$ and $\lambda_k \geqslant 0$ is defined as in step [3.2] of algorithm 1;
(c) $h_k = x_{k+1} - x_k = 0$ if and only if $A_i x_k = y_i$ (see step [3.2] of algorithm 1); in this case, the unique successor of x_k is $z = x_k$ itself.

Theorem 4.3 (Stability). *Assume that (A1) and (A2) hold true. Let $\delta^j := (\delta_0^j, \ldots, \delta_{N-1}^j)_{j\in\mathbb{N}}$ be a zero sequence and $\mathbf{y}^{\delta^j} = (y_0^{\delta^j}, \ldots, y_{N-1}^{\delta^j})_{j\in\mathbb{N}}$ a corresponding sequence of noisy data satisfying (1). For each $j \in \mathbb{N}$, let $x_{k+1}^{\delta^j}$ be a successor of $x_k^{\delta^j}$ for $0 \leqslant k \leqslant k_*(\delta^j, \mathbf{y}^{\delta^j})$. Then, there exists a noiseless sequence $(x_k)_{k\in\mathbb{N}}$ such that, for every fixed $k \in \mathbb{N}$, there exists a subsequence (δ^{j_m}), which depends on k, satisfying*

$$x_l^{\delta^{j_m}} \to x_l, \quad \textit{as } j_m \to \infty, \textit{ for } l = 0, \ldots, k.$$

Proof. We use an inductive argument. Since $x_0^\delta = x_0$ for every $\delta \geqslant 0$, the assertion is clear for $k = 0$. Our main argument consists of repeatedly choosing a subsequence of the current subsequence. In order to avoid a notational overload, we denote a subsequence of $(\delta^j)_j$ again by $(\delta^j)_j$. Suppose by induction that the assertion holds true for some $k \in \mathbb{N}$, i.e., that there exists a subsequence $(\delta^j)_j$ and $(x_l)_{l=0}^k$ satisfying

$$x_l^{\delta j} \to x_l, \quad \text{as } j \to \infty, \text{ for } l = 0, \ldots, k, \tag{20}$$

where x_{l+1} is a successor of x_l for $l = 0, \ldots, k-1$.[5]

Since $x_{k+1}^{\delta j}$ is a successor of $x_k^{\delta j}$ (for each δ^j) there exists a positive number $\lambda_k^{\delta j}$ such that $x_{k+1}^{\delta j} = x_k^{\delta j} - \lambda_k^{\delta j}\left(I + \lambda_k^{\delta j} A_i^* A_i\right)^{-1} A_i^* \left(A_i\, x_k^{\delta j} - y_i^{\delta j}\right)$. Our next goal is to prove the existence of a successor x_{k+1} of x_k, and of a subsequence $(\delta^j)_j$ of the current sequence, such that $x_{k+1}^{\delta j} \to x_{k+1}$ as $j \to \infty$, completing the inductive proof. We divide this proof in three steps as follows:

Step 1. We define an element $x_{k+1} \in X$;

Step 2. We show that

$$x_{k+1}^{\delta j} \to x_{k+1}, \quad \text{as } j \to \infty; \tag{21}$$

Step 3. We prove that x_{k+1} is a successor of x_k.

Proof of step 1. It follows from algorithm 1 that $\lambda_k^{\delta_j} \leqslant \lambda_{\max}$ for each $k \in \mathbb{N}$ fixed (see remark 2.2). Consequently, there exists a $\lambda_k > 0$ satisfying

$$\lambda_k := \lim_{j\to\infty} \lambda_k^{\delta_j} < \infty. \tag{22}$$

If $\lambda_k = 0$ we set $x_{k+1} := x_k$, otherwise we set $x_{k+1} := \arg\min_{x\in X} T_{k,\lambda_k}(x)$, with T_{k,λ_k} as in definition 4.2 item 2.

Proof of step 2. From definition of $x_{k+1}^{\delta j}$ and x_{k+1} it follows that

$$\begin{aligned} \lambda_k^{\delta j} A_i^*(A_i x_{k+1}^{\delta j} - y_i^{\delta j}) + x_{k+1}^{\delta j} - x_k^{\delta j} &= 0 \\ &= \lambda_k A_i^*(A_i x_{k+1} - y_i) + x_{k+1} - x_k, \end{aligned}$$

from where we obtain

$$\begin{aligned} 0 = {}& \lambda_k^{\delta j} A_i^* A_i (x_{k+1}^{\delta j} - x_{k+1}) + x_{k+1}^{\delta j} - x_{k+1} + (\lambda_k^{\delta j} - \lambda_k) A_i^* A_i x_{k+1} \\ & + \lambda_k y_i - \lambda^{\delta j} y_i^{\delta j} - x_k^{\delta j} + x_k. \end{aligned}$$

Multiplying the last expression by $x_{k+1}^{\delta j} - x_{k+1}$ we get

$$\begin{aligned} \|x_{k+1}^{\delta j} - x_{k+1}\|^2 &\leqslant \lambda_k^{\delta j} \|A_i(x_{k+1}^{\delta j} - x_{k+1})\|^2 + \|x_{k+1}^{\delta j} - x_{k+1}\|^2 \\ &= |\lambda_k - \lambda_k^{\delta j}| \langle A_i^* A_i x_{k+1}, (x_{k+1}^{\delta j} - x_{k+1}) \rangle \\ &\quad + \langle x_k - x_k^{\delta j}, x_{k+1}^{\delta j} - x_{k+1} \rangle + \langle \lambda_k^{\delta j} y_i^{\delta j} - \lambda_k y_i, x_{k+1}^{\delta j} - x_{k+1} \rangle \\ &\leqslant C^2 |\lambda_k - \lambda_k^{\delta j}| \, \|x_{k+1}\| \, \|x_{k+1}^{\delta j} - x_{k+1}\| + \|x_k - x_k^{\delta j}\| \\ &\quad \times \|x_{k+1}^{\delta j} - x_{k+1}\| + \|\lambda_k y_i - \lambda_k^{\delta j} y_i^{\delta j}\| \, \|x_{k+1}^{\delta j} - x_{k+1}\|. \end{aligned}$$

[5] Notice that $k_*(\delta^j, \mathbf{y}^{\delta^j}) \geqslant k$ for large enough j.

Consequently $\|x_{k+1}^{\delta j} - x_{k+1}\| \leqslant C^2|\lambda_k - \lambda_k^{\delta j}|\,\|x_{k+1}\| + \|x_k - x_k^{\delta j}\| + \|\lambda_k y_i - \lambda_k^{\delta j} y_i^{\delta j}\|$. In view of (1), (20) and (22) it follows that (21) holds.

Proof of step 3. We consider two cases:

First case: $\|A_i x_k - y_i\| = 0$: in order to prove that x_{k+1} is a successor of x_k, it is enough to prove that $x_{k+1} = x_k$ (see definition 4.2). From the definition of x_{k+1} in step 1 we have either $x_{k+1} = x_k$ (and we are done) or $x_{k+1} := \arg\min_{x\in X} T_{k,\lambda_k}(x)$. From where we obtain $\lambda_k A_i^*[A_i(x_{k+1} - x_k) + A_i x_k - y_i] + (x_{k+1} - x_k) = 0$. Thus, $(\lambda_k A_i^* A_i + I)(x_{k+1} - x_k) = 0$ and $x_{k+1} = x_k$ follows.

Second case: $\|A_i x_k - y_i\| > 0$: it follows from (20) that $\lim_j \|A_i x_k^{\delta j} - y_i^{\delta j}\| = \|A_i x_k - y_i\| > 0$. Consequently, for sufficiently large j we have $\|A_i x_k^{\delta j} - y_i^{\delta j}\| > \tau\delta_i^j$,

$$x_{k+1}^{\delta j} = x_k^{\delta j} - \lambda_k^{\delta j}\left(I + \lambda_k^{\delta j} A_i^* A_i\right)^{-1} A_i^*\left(A_i\, x_k^{\delta j} - y_i^{\delta j}\right) \tag{23}$$

and

$$\bar{p}\|A_i x_k^{\delta j} - y_i^{\delta j}\| + (1-\bar{p})\delta_i^j \leqslant \|A_i x_{k+1}^{\delta j} - y_i^{\delta j}\| \leqslant \bar{\bar{p}}\|A_i x_k^{\delta j} - y_i^{\delta j}\| + (1-\bar{\bar{p}})\delta_i^j. \tag{24}$$

Since $x_{k+1}^{\delta j} \to x_{k+1}$ as $j \to \infty$ (see step 2.), we take the limit $j \to \infty$ in (23) and (24) to conclude that (λ_k, h_k) (with $h_k := x_{k+1} - x_k$, defined in step 1) satisfy the range relaxed problem in step [3.2] of algorithm 1. Consequently, x_{k+1} is a successor of x_k. □

Theorem 4.4 (Semi-convergence). *Assume that (A1) and (A2) hold true. Let $\delta^j := (\delta_0^j, \ldots, \delta_{N-1}^j)_{j\in\mathbb{N}}$ be a zero sequence and $\mathbf{y}^{\delta j} = (y_0^{\delta j}, \ldots, y_{N-1}^{\delta j})_{j\in\mathbb{N}}$ a corresponding sequence of noisy data satisfying (1). For each $j \in \mathbb{N}$, let $x_{k+1}^{\delta j}$ be a successor of $x_k^{\delta j}$ for $0 \leqslant k \leqslant k_*^j = k_*(\delta^j, \mathbf{y}^{\delta j})$. Then, every subsequence of $x_{k_*^j}^{\delta j}$ has itself a subsequence converging strongly to $x^\dagger$, the x_0-minimal norm solution of (2).*

Proof. We consider two cases: assume first that the sequence $(k_*^j)_{j\in\mathbb{N}}$ is bounded. Then it has a finite accumulation point. In this case, since $(k_*^j) \in \mathbb{N}$ we can extract a subsequence $(\delta^{j_m})_{m\in\mathbb{N}}$ of (δ^j) such that $k_*^{j_m} = n$ for some $n \in \mathbb{N}$, and all j_m. From theorem 4.3, the subsequence $(x_n^{\delta j_m})_{m\in\mathbb{N}}$ has itself a subsequence (denoted again by $(x_n^{\delta j_m})_{m\in\mathbb{N}}$) converging to x_n, i.e.,

$$\lim_{m\to\infty} x_{k_*^{j_m}}^{\delta j_m} = \lim_{m\to\infty} x_n^{\delta j_m} = x_n.$$

Notice that x_n is a solution of (2). Indeed, for $i \in \{0, \ldots, N-1\}$ we have

$$\begin{aligned}\|A_i x_n - y_i\| &= \lim_{m\to\infty}\|A_i x_{k_*^{j_m}}^{\delta j_m} - y_i\|\\ &\leqslant \lim_{m\to\infty}\left(\|A_i x_{k_*^{j_m}}^{\delta j_m} - y_i^{\delta j_m}\| + \|y_i^{\delta j_m} - y_i\|\right)\\ &\leqslant \lim_{m\to\infty}(\tau+1)\delta^{j_m} = 0.\end{aligned}$$

In the second case we assume that $(k_*^j)_j$ is not bounded. Thus there exists a monotone increasing subsequence, again denoted by (k_*^j), such that $k_*^j \to \infty$ as $j \to \infty$. Fix $\varepsilon > 0$ and let $(x_k)_{k\in\mathbb{N}}$ be a noiseless sequence (see definition 4.2). From theorem 3.2 follows that $x_k \to x^\dagger$, where $x^\dagger$ is the x_0-minimum norm solution of (2). Then there exists an index $L = L(\varepsilon) \in \mathbb{N}$ such that,

$$\|x_k - x^\dagger\| < \frac{\varepsilon}{2}, \quad k \geqslant L. \tag{25}$$

Moreover, since $k_*^j \to \infty$ as $j \to \infty$, there exists $J \in \mathbb{N}$ such that $k_*^j \geqslant L$, for all $j \geqslant J$. Furthermore, from corollary 2.5 follows

$$j \geqslant J \;\Rightarrow\; \left\| x_{k_*^j}^{\delta j} - x^\dagger \right\| \leqslant \left\| x_L^{\delta j} - x^\dagger \right\|,$$

and from theorem 4.3 follows the existence of a subsequence (δ^{j_m}) (depending on $L(\varepsilon)$) and of an $M \in \mathbb{N}$ such that

$$m \geqslant M \implies \left\| x_L^{\delta j_m} - x_L \right\| < \frac{\varepsilon}{2}.$$

Thus, for $j_m \geqslant \max\{J, M\}$ we have

$$\left\| x_{k_*^{j_m}}^{\delta j_m} - x^\dagger \right\| \leqslant \left\| x_L^{\delta j_m} - x^\dagger \right\| \leqslant \left\| x_L^{\delta j_m} - x_L \right\| + \|x_L - x^\dagger\| < \varepsilon. \tag{26}$$

Note that the subsequence (δ^{j_m}) depends on ε. We construct an independent subsequence using a diagonal argument: choosing $\varepsilon = 1$, we can find a subsequence $(\delta^j)_{j\in\mathbb{N}}$ and select a number $j_1 \in \mathbb{N}$ such that

$$\left\| x_{k_*^{j_1}}^{\delta j_1} - x^\dagger \right\| < 1. \tag{27}$$

Since the current subsequence $(\delta^j)_{j\in\mathbb{N}}$ is also a positive-zero sequence, the above reasoning can be applied again with $\varepsilon = \frac{1}{2}$. We choose a number $j_2 \geqslant j_1$ such that

$$\left\| x_{k_*^{j_2}}^{\delta j_2} - x^\dagger \right\| < \frac{1}{2}. \tag{28}$$

Using induction, it is therefore possible to construct a subsequence $(\delta^{j_m})_{m\in\mathbb{N}}$ with the property

$$\left\| x_{k_*^{j_m}}^{\delta j_m} - x^\dagger \right\| < \frac{1}{m}, \quad \text{for all } m \in \mathbb{N}, \tag{29}$$

this is implies that $\left\| x_{k_*^{j_m}}^{\delta j_m} - x^\dagger \right\| \to 0$, as $m \to \infty$. □

5. Numerical experiments

In this section the rriTK method, algorithm 1, is implemented for solving two distinct inverse problems, namely: 1) the image deblurring problem [4]; 2) the IPP [22].

The performance of the rriTK method is compared against three other Kaczmarz type methods, namely: the Landweber Kaczmarz (LWK), the geometric iterated Tikhonov Kaczmarz (giTK) with $\lambda_k = 2^k$ and the stationary iterated Tikhonov Kaczmarz method (siTK) with $\lambda_k = 2$ methods.

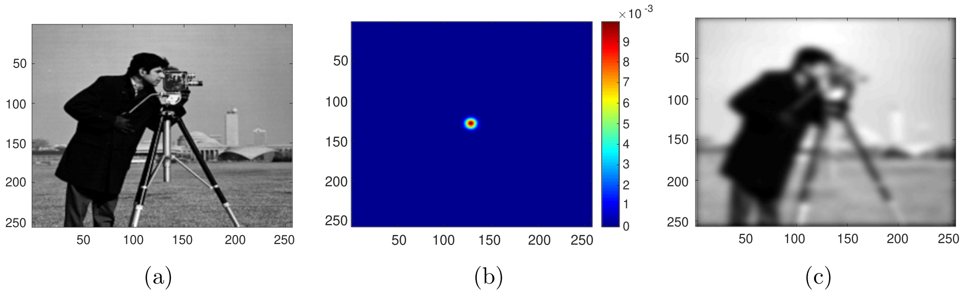


Figure 1. Image deblurring: setup of the inverse problem. (a) Original image x; (b) point spread function; (c) blurred image y.

Table 1. Image deblurring: number of computed cycles; in parentheses the number of computed steps.

δ	rriTK	giTK
10%	15 (124)	111 (850)
1%	24 (212)	116 (939)
0.1%	29 (262)	119 (1023)

5.1. Image deblurring

These are finite dimensional problems modeled, in general, by high dimensional linear systems of the form $Ax = y$. Here the vector $x \in X = \mathbb{R}^n$ represents the pixel values of an unknown true image, while the vector $y \in Y = X$ contains the pixel values of the observed (blurred) image. The operator $A : X \to X$ describes the (discretized) blurring phenomenon [3, 4]. We consider the situation where the blur of the image is modeled by a space invariant point spread function (PSF).

In the continuous setting, the blurring process is represented by an integral operator of convolution type. Thus, the mathematical model corresponds to an integral equation of the first kind [13]. In the discrete setting, after incorporating appropriate boundary conditions into the model, the discrete convolution is evaluated by means of the FFT algorithm.

The setup of our deblurring experiment is shown in figure 1: (a) True image $x \in \mathbb{R}^n$, $n = 256^2$ (Cameraman 256×256); (b) PSF is the rotationally symmetric Gaussian low-pass filter of size [256 256] and standard deviation $\sigma = 4$; (c) Exact data $y = Ax \in \mathbb{R}^n$ (blurred image).

The linear operators A_i in (2) correspond to blocks of lines of the (discretized) blurring operator A (i.e., $N = 16$ blocks with 32 lines each), while the data y_i is defined accordingly. In the implementation of the rriTK method we use the values $\bar{p} = 0.1$, $\bar{\bar{p}} = 0.5$. All implemented methods are stopped according to the discrepancy principle with $\tau = 1.5$. The available data y_i^δ is generated by adding artificial (white) noise to the blurred image y_i. As initial (for all methods) guess we set $x_0 = y^\delta$ (the noisy blurred image).

In our experiments, relative noise $\|y_i - y_i^\delta\|/\|y_i\|$ of three distinct levels, namely 0.1%, 1% and 10%, are used. The results are summarized in table 1, where the number of computed cycles for each method is shown, as well as the number of computed steps. It is worth mentioning that the LWK and siTK methods have not reached the stop criteria (these methods were stopped after 10^4 cycles).

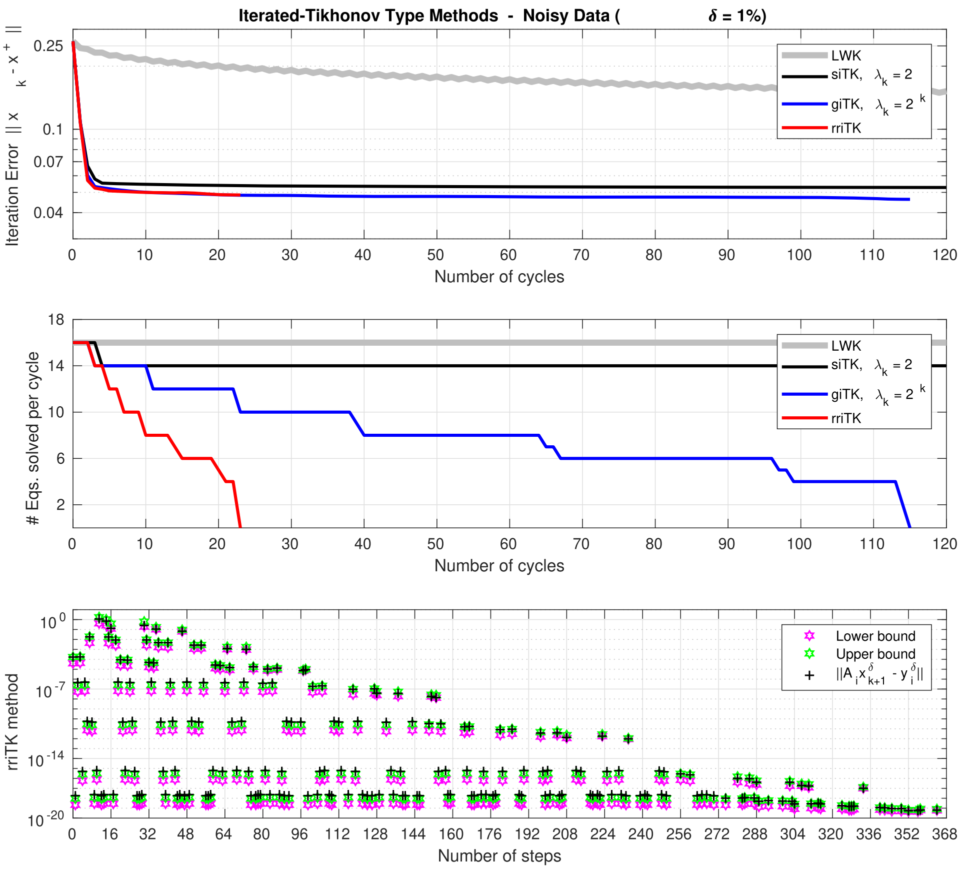


Figure 2. Image deblurring: noisy data case $\delta = 1\%$. (TOP) Iteration error; (CENTER) number of equations solved per cycle; (BOTTOM) verification of the inequalities $\bar{\Phi} \leqslant \|b_k^\delta\| \leqslant \bar{\bar{\Phi}}$ in step [3.2].

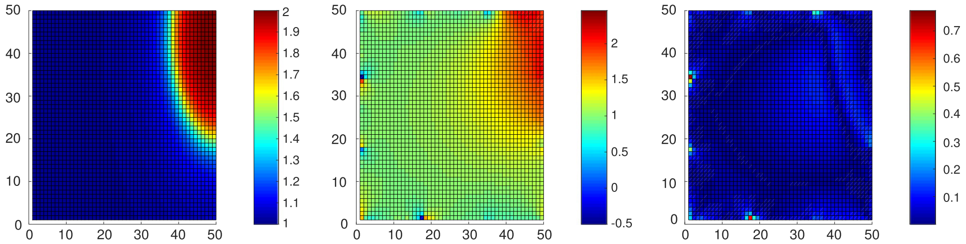
A more detailed comparison is presented in figure 2 for the noise level of $\delta_i = 1\%$: (GRAY) LWK method; (BLACK) siTK with $\lambda_k = 2$; (BLACK) giTK with $\lambda_k = 2^k$; (RED) rriTK method (with $\bar{p} = 0.1$ and $\bar{\bar{p}} = 0.5$).

The pictures in figure 2 show: (TOP) relative error $\|x^\star - x_k^\delta\|/\|x^\star\|$; (CENTER) number of equations solved in each cycle; (BOTTOM) verification of the inequalities $\bar{\Phi} \leqslant \|b_k^\delta\| \leqslant \bar{\bar{\Phi}}$ in step [3.2] of algorithm 1.

The x-axis in the (TOP) and (CENTER) pictures is scaled by the number of cycles. In the (BOTTOM) picture, the x-axis shows the number of steps (notice that each cycle consists of 16 iterative steps).

5.2. Inverse potential problem

The IPP is a parameter identification problem for elliptic PDE's [8, 15, 22, 34]. Generalizations of this inverse problem appear in many relevant applications including inverse gravimetry [24, 34], EEG [12], and EMG [35].



(a) (b) (c)

Figure 3. IPP: second noise scenario $\delta = 0.1\%$. (a) Exact solution $x^\star$; (b) approximate solution x^δ_{72} (rriTK method); (c) iteration error $|x^\star - x^\delta_{72}|$ (absolute error pixel-wise).

Table 2. IPP: number of computed cycles; in parentheses the number of computed steps.

δ	rriTK	giTK	siTK	LWK
1%	2 (10)	3 (21)	6 (32)	10 (56)
0.1%	6 (43)	5 (55)	20 (165)	35 (298)
0.025%	7 (64)	7 (73)	44 (358)	88 (669)

The forward problem consists in solving on a Lipschitz domain $\Omega \subset \mathbb{R}^d$, for a given source function $x \in L_2(\Omega)$, the boundary value problem

$$-\Delta u \;=\; x, \quad \text{in } \Omega, \qquad u \;=\; 0 \quad \text{on } \partial\Omega. \tag{30}$$

The corresponding inverse problem is the so called IPP, which consists of recovering an L_2-function x, from measurements of the Neumann data of its corresponding H^1-potential u on the boundary of Ω, i.e., $y := u_\nu|_{\partial\Omega} \in L_2(\partial\Omega)$.

The IPP problem is modeled by the linear operator $A : L_2(\Omega) \to L_2(\partial\Omega)$ defined by $Ax := u_\nu|_{\partial\Omega}$, where $u \in H^1_0(\Omega)$ is the unique solution of (30) (see [22]). Using this notation, the IPP can be written in the abbreviated form $Ax = y$.

In our experiments we follow [5] in the experimental setup, selecting $\Omega = (0, 1) \times (0, 1)$ and assuming that the unknown parameter $x^\star$ is an H^1-function with sharp gradients (see figure 3(a)). In the discrete setting, the solution of the involved elliptic BVP's is computed using finite differences on a uniform mesh with 50^2 nodes. Thus the operator A is approximated by a matrix $A_d : \mathbb{R}^{2500} \to \mathbb{R}^{192}$. The boundary of Ω is divided in $N = 12$ segments, i.e. $\partial\Omega = \cup_{i=0}^{11}\Gamma_i$, and each Γ_i is discretized using 16 boundary nodes. Thus, the linear operators in (2) are defined by $A_i := \gamma_i \circ A_d$, where $\gamma_i : \mathbb{R}^{192} \to \mathbb{R}^{16}$ is the corresponding discretization of the projection operator from $\partial\Omega$ to Γ_i. The (discretized) boundary data $y_i = \gamma_i(y)$ are defined accordingly.

Perturbed data y_i^δ are generated by adding to the exact Neuman data y_i a normally distributed noise with zero mean and suitable variance for achieving a prescribed relative noise level. In the numerical implementations we set $\bar{p} = 0.1$, $\bar{\bar{p}} = 0.5$, $\tau = 2$ (discrepancy principle constant) and use the initial guess $x_0 \equiv 1.5$ (constant function in Ω).

As in section 5.1, three distinct scenarios are considered, where the relative noise level $\|y - y^\delta\|/\|y\|$ corresponds to 1.0%, 0.1% and 0.025% respectively. The results are summarized in table 2.

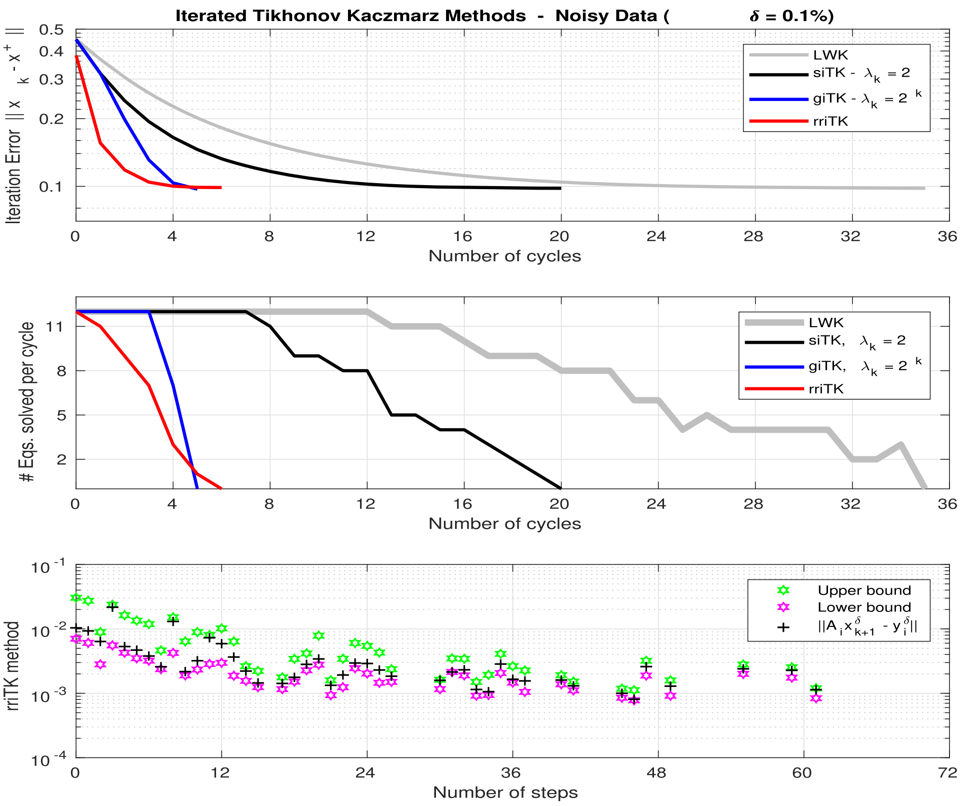


Figure 4. Inverse Potential Problem: noisy data case $\delta = 0.1\%$. (TOP) Iteration error; (CENTER) number of equations solved per cycle; (BOTTOM) verification of the inequalities $\bar{\Phi} \leqslant \|b_k^\delta\| \leqslant \bar{\bar{\Phi}}$.

In figure 4 the following methods are compared for the second noise scenario $\delta = 0.1\%$: (GRAY) LWK method; (BLACK) siTK with $\lambda_k = 2$; (BLACK) giTK with $\lambda_k = 2^k$; (RED) rriTK method (with $\bar{p} = 0.1$ and $\bar{\bar{p}} = 0.5$). The pictures in this figure show: (TOP) relative error $\|x^\star - x_k^\delta\|/\|x^\star\|$; (CENTER) number of equations solved in each cycle; (BOTTOM) verification of the inequalities $\bar{\Phi} \leqslant \|b_k^\delta\| \leqslant \bar{\bar{\Phi}}$ in step [3.2] of algorithm 1.

For the noise scenario $\delta = 0.1\%$, the iterate x_{72}^δ (computed by rriTK after 6 cycles) and the corresponding iteration error $|x^\star - x_{72}^\delta|$ are shown in figures 3(b) and (c) respectively.

5.3. Remarks on the numerical experiments

The computation of our numerical experiments were conducted using MATLAB 2012a. Direct and inverse problems were solved using different levels of discretization in order to avoid inverse crimes.

For the IPPs, in all noise scenarios, both rriTK and giTK require similar number of cycles to reach the same stop criteria[6]. On the other hand, for the Deblurring problem, rriTK requires

[6] In the giTK we chose λ_k constant within the cycles, i.e., $\lambda_k = 2^{\kappa+1}$ with $\kappa = k \backslash N$ (integer division).

less cycles than giTK. It is worth noticing that, for both inverse problems and all noise scenarios rriTK computes less steps than the giTK (compare the numbers in parentheses in tables 1 and 2).

In the numerical implementations of algorithm 1 for both inverse problems above, we controlled the inequalities in step [3.2] for all computed steps. This can be verified in the (BOTTOM) pictures of figures 2 and 4. In these pictures one can also recognize which steps were actually computed (i.e., $h_k \neq 0$) in each of the cycles of the rriTK method.

6. Conclusions and future work

We investigate nonstationary iTK type methods for computing stable approximate solutions to systems of linear ill-posed operator equations. The main contribution of this article is to extend the strategy for choosing the Lagrange multipliers for the iT method in [5] in order to couple the iT method with the Kaczmarz strategy (we propose a different range as the one in [5]). This modification allows us to prove convergence for exact data (section 3) using a technique completely different from the one applied in [5]. Moreover, we also prove stability and semi-convergence results (section 4), which are not considered in [5].

The range-relaxed strategy is very advantageous, since it allows each of the multipliers to belong to a non-degenerate interval. Consequently, the actual computation of the Lagrange multipliers (satisfying the theoretical requirements needed for the convergence analysis) is simplified.

An algorithmic implementation of the rriTK method is discussed (algorithm 1), and it is tested for solving two well known ill-posed problems (Image Deblurring and IPP) using three different levels of noise.

The numerical experiments in section 5 show that the rriTK method is competitive with other Kaczmarz type methods, including the giTK method, where the Lagrange multipliers are computed *a priori* in geometric progression (a commonly used strategy).

The careful reader observes that many of the references cited in this manuscript relate to methods for solving non-linear systems of equations. A natural question arises: what are the difficulties to extend the approach presented in this paper to this non-linear framework? So far, we have only a partial answer to that question. Using a nonlinear assumption called *tangential cone condition* we are able to extend the 'gain inequality' for a nonlinear version of the rriTK method (namely, the rrLMK or range-relaxed Levenberg Marquardt Kaczmarz method). In a future work, we aim to extend the convergence analysis derived in this paper to the rrLMK method for solving systems of nonlinear ill-posed equations.

Acknowledgments

JCR acknowledges support from the Federal University of Piaui. AL acknowledges support from the research agency CNPq (Grant No. 311087/2017-5), and from the AvH Foundation.

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 (Dordrecht: Springer)
[2] Baumeister J 1987 *Stable Solution of Inverse Problems* (*Advanced Lectures in Mathematics*) (Braunschweig: Vieweg)
[3] Bertero M, Boccacci P, Desiderà G and Vicidomini G 2009 Image deblurring with Poisson data: from cells to galaxies *Inverse Problems* **25** 123006
[4] Bertero M and Boccacci P 1998 *Introduction to Inverse Problems in Imaging* (*Advanced Lectures in Mathematics*) (Bristol: Institute of Physics Publishing)
[5] Boiger R, Leitão A and Svaiter B F 2020 Range-relaxed criteria for choosing the Lagrange multipliers in nonstationary iterated Tikhonov method *IMA J. Numer. Anal.* **40** 606–27
[6] Brill M and Schock E 1987 Iterative solution of ill-posed problems: a survey *Model Optimization in Exploration Geophysics* ed A Vogel (Braunschweig: Vieweg) pp 13–38
[7] Burger M and Kaltenbacher B 2006 Regularizing Newton–Kaczmarz methods for nonlinear ill-posed problems *SIAM J. Numer. Anal.* **44** 153–82
[8] De Cezaro A, Leitão A and Tai X-C 2009 On multiple level-set regularization methods for inverse problems *Inverse Problems* **25** 035004
[9] De Cezaro A, Baumeister J and Leitão A 2011 Modified iterated Tikhonov methods for solving systems of nonlinear ill-posed equations *Inverse Problems Imaging* **5** 1–17
[10] 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
[11] Donatelli M 2012 On nondecreasing sequences of regularization parameters for nonstationary iterated Tikhonov *Numer. Algorithm.* **60** 651–68
[12] El Badia A and Farah M 2006 Identification of dipole sources in an elliptic equation from boundary measurements: application to the inverse EEG problem *J. Inverse Ill-Posed Problems* **14** 331–53
[13] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer Academic Publishers)
[14] Fakeev A G 1981 A class of iterative processes for solving degenerate systems of linear algebraic equations *USSR Comput. Math. Math. Phys.* **21** 15–22
[15] Frühauf F, Scherzer O and Leitão A 2005 Analysis of regularization methods for the solution of ill-posed problems involving discontinuous operators *SIAM J. Numer. Anal.* **43** 767–86
[16] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (*Research Notes in Mathematics*) vol 105 (Boston, MA: Pitman (Advanced Publishing Program))
[17] Haltmeier M, Kowar R, Leitão A and Scherzer O 2007 Kaczmarz methods for regularizing nonlinear ill-posed equations. II. Applications *Inverse Problems Imaging* **1** 507–23
[18] Haltmeier M, Leitão A and Scherzer O 2007 Kaczmarz methods for regularizing nonlinear ill-posed equations I: convergence analysis *Inverse Problems Imaging* **1** 289–98
[19] Haltmeier M, Leitão A and Resmerita E 2009 On regularization methods of EM-Kaczmarz type *Inverse Problems* **25** 075008
[20] Hanke M and Groetsch C W 1998 Nonstationary iterated Tikhonov regularization *J. Optim. Theory Appl.* **98** 37–53
[21] 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
[22] Hettlich F and Rundell W 1996 Iterative methods for the reconstruction of an inverse potential problem *Inverse Problems* **12** 251–66
[23] Hofmann B 1986 Regularization for applied inverse and ill-posed problems *Teubner-Texte zur Mathematik* (*Teubner Texts in Mathematics*) vol 85 (Leipzig: BSB B G Teubner Verlagsgesellschaft) A numerical approach, With German, French and Russian summaries
[24] Isakov V 2006 *Inverse Problems for Partial Differential Equations* (*Applied Mathematical Sciences*) 2nd edn vol 127 (New York: Springer)
[25] Kaczmarz S 1993 Approximate solution of systems of linear equations *Int. J. Control* **57** 1269–71
[26] 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: Walter de Gruyter GmbH & Co. KG)
[27] Kirsch A 1996 *An Introduction to the Mathematical Theory of Inverse Problems* (*Applied Mathematical Sciences*) vol 120 (New York: Springer)

[28] Lardy L J 1975 A series representation for the generalized inverse of a closed linear operator *Atti della Accademia Nazionale dei Lincei, Rendiconti della Classe di Scienze Fisiche, Matematiche, e Naturali, Serie VIII* **58** 152–7
[29] Leitão A, Margotti F and Svaiter B F 2020 Range-relaxed criteria for choosing the Lagrange multipliers in the Levenberg-Marquardt method *IMA J. Numer. Anal.* (unpublished)
[30] Louis A 1989 *Inverse und schlecht gestellte Probleme* (Stuttgart: B G Teubner)
[31] Morozov V A 1993 *Regularization Methods for Ill–Posed Problems* (Boca Raton, FL: CRC Press)
[32] Natterer F 2001 *The Mathematics of Computerized Tomography* (Philadelphia, PA: SIAM)
[33] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (Philadelphia, PA: SIAM)
[34] van den Doel K, Ascher U M and Leitão A 2010 Multiple level sets for piecewise constant surface reconstruction in highly ill-posed problems *J. Sci. Comput.* **43** 44–66
[35] van den Doel K, Ascher U M and Pai D K 2008 Computed myography: three-dimensional reconstruction of motor functions from surface EMG data *Inverse Problems* **24** 065010