Inverse Problems **25** (2009) 075008 (17pp) doi:10.1088/0266-5611/25/7/075008

On regularization methods of EM-Kaczmarz type

M Haltmeier[1], **A Leitão**[2] **and E Resmerita**[3]

[1] Department of Mathematics, University of Innsbruck, Technikerstraße 21a, A-6020 Innsbruck, Austria
[2] Department of Mathematics, Federal University of St. Catarina, PO Box 476, 88040-900 Florianopolis, Brazil
[3] Industrial Mathematics Institute, Johannes Kepler University, Altenbergerstraße 69, A-4040 Linz, Austria

E-mail: markus.haltmeier@uibk.ac.at, acgleitao@gmail.com and elena.resmerita@jku.at

Received 1 March 2009, in final form 11 May 2009
Published 12 June 2009
Online at stacks.iop.org/IP/25/075008

Abstract

We consider regularization methods of Kaczmarz type in connection with the expectation-maximization (EM) algorithm for solving ill-posed equations. For noisy data, our methods are stabilized extensions of the well-established ordered-subsets expectation-maximization iteration (OS-EM). We show monotonicity properties of the methods and present a numerical experiment which indicates that the extended OS-EM methods we propose are much faster than the standard EM algorithm in a relevant application.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The *expectation-maximization* (EM) algorithm introduced in [3] provides approximations for maximum likelihood estimators of problems with incomplete or noisy data, which is the usual framework when dealing with inverse or ill-posed problems. In particular, the EM algorithm for Poisson models is well known for its applications to astronomical imaging and to positron emission tomography (PET)—see, e.g., [27, 29].

In this work we address inverse problems modeled by operator equations with nonnegative data and which admit nonnegative solutions, with the aim of approaching them by combined EM-Kaczmarz strategies.

We begin our study by considering the operator equation

$$\mathcal{A}x = y, \tag{1.1}$$

where $\mathcal{A} : L^1(\Omega) \to L^1(\Sigma)$ is a Fredholm integral operator of the first kind

$$(\mathcal{A}x)(s) = \int_\Omega a(s,t)x(t)\,\mathrm{d}t, \qquad s \in \Sigma. \tag{1.2}$$

The EM algorithm for solving the equation reads as follows:

$$x_{k+1}(t) = x_k(t)\mathcal{A}^*\left(\frac{y}{\mathcal{A}x_k}\right)(t) = x_k(t)\int_\Sigma \frac{a(s,t)y(s)}{(\mathcal{A}x_k)(s)}\mathrm{d}s. \tag{1.3}$$

This is the continuous version (see, e.g., [5, 15, 19–22, 26]) of the celebrated EM algorithm for PET (see, e.g., [8, 13, 18, 24, 29]).

Note that, nonnegative solutions of (1.1) can be determined by finding minimizers of the functional

$$f(x) := \int_\Sigma \left[y(s)\ln\frac{y(s)}{(\mathcal{A}x)(s)} - y(s) + (\mathcal{A}x)(s)\right]\mathrm{d}s.$$

Formally, the first-order necessary condition for such a minimizer is

$$\mathcal{A}^*\left(\frac{y}{\mathcal{A}x}\right) = \mathcal{A}^*1, \tag{1.4}$$

where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$. If the assumption $(\mathcal{A}^*1)(t) = 1$ is satisfied, a solution to (1.4) can be obtained by solving the corresponding multiplicative fixed-point equation

$$x\mathcal{A}^*\left(\frac{y}{\mathcal{A}x}\right) = x. \tag{1.5}$$

Thus, one can see that the minimizers of the function f are strongly related to the fixed-point equation (1.5) and, therefore, to the EM algorithm (1.3).

The ordered-subsets expectation-maximization (OS-EM) iteration was introduced in [12] as a computationally more efficient alternative to the original EM iteration for the discrete case. The main idea is as follows. The data y are grouped into an ordered sequence of subsets (or blocks) y_j. An iteration of OS-EM consists of a single cycle through all the subsets, in each subset updating the current estimate by an application of the EM algorithm in that data subset. This strategy can be connected to the Kaczmarz-type iterative methods recently investigated in [2, 9, 10, 16] for approaching systems of integral equations.

In order to extend the OS-EM method to infinite-dimensional settings, we first group the data y into N blocks $y_j := y|_{\Sigma_j}$, where $\Sigma_j \subset \Sigma$ are not necessarily disjoint and satisfy $\Sigma = \Sigma_0 \cup \cdots \cup \Sigma_{N-1}$. Then equation (1.1) is decomposed into a system of operator equations of the first kind

$$\mathcal{A}_j x = y_j, \qquad j = 0, \ldots, N-1, \tag{1.6}$$

where the Fredholm integral operators $\mathcal{A}_j : L^1(\Omega) \to L^1(\Sigma_j)$ correspond to blocks of $\mathcal{A}$ and are defined by

$$(\mathcal{A}_j x)(s) := \int_\Omega a_j(s,t)x(t)\,\mathrm{d}t, \tag{1.7}$$

with $a_j := a|_{\Omega\times\Sigma_j}$. Note that x is a solution to (1.6) if and only if x solves (1.1).

Without loss of generality, we drop the indices of the domains Σ_j and write $\mathcal{A}_j : L^1(\Omega) \to L^1(\Sigma)$ and $y_j \in L^1(\Sigma)$. Thus, the system of integral equations (1.6) can be approached by simultaneously minimizing

$$f_j(x) := \int_\Sigma \left[y_j(s)\ln\frac{y_j(s)}{(\mathcal{A}_j x)(s)} - y_j(s) + (\mathcal{A}_j x)(s)\right]\mathrm{d}s = d(y_j, \mathcal{A}_j x), \tag{1.8}$$

where $d(u, v)$ is the Kullback–Leibler (KL) distance defined by

$$d(v, u) := \int \left[v(t)\ln\frac{v(t)}{u(t)} - v(t) + u(t)\right]\mathrm{d}t. \tag{1.9}$$

Throughout this paper we will make use of the KL-distance $d(v, u)$ with either $u, v \in L^1(\Omega)$ or $u, v \in L^1(\Sigma)$.

Remark 1.1. Analog as in (1.4), if the assumption $\mathcal{A}_j^*1 = 1$, $j = 0, \ldots, N-1$, is satisfied, then the first-order necessary condition for a minimizer of f_j is given by $\mathcal{A}_j^*(y_j/(\mathcal{A}_j x)) = 1$, and the corresponding multiplicative fixed-point equation reads $P_j(x) := x\mathcal{A}_j^*(y_j/(\mathcal{A}_j x)) = x$.

The OS-EM algorithm corresponds to a Kaczmarz-type method for solving system (1.6) and can be written in the form

$$x_{k+1} = P_j(x_k) = x_k \int_\Sigma \frac{a_j(s,\cdot)y_j(s)}{(\mathcal{A}_j x_k)(s)}\,\mathrm{d}s, \tag{1.10}$$

where the index $0 \leqslant j < N$ relates to the iteration index k by the formula $j = [k] := (k \bmod N)$. Clearly, the case $N = 1$ corresponds to the standard EM algorithm.

The cyclic structure of the iteration in (1.10) is easily recognizable (each cycle consists of N steps). Note that each step within a cycle is an explicit step for solving the fixed point equation $x\mathcal{A}_{[k]}^*(y_{[k]}/(\mathcal{A}_{[k]}x)) = x$, and can be interpreted as an EM iterative step for solving the $[k]$th equation (or block) of system (1.6).

This paper is outlined as follows. In section 2, we formulate a series of assumptions, which are necessary for the analytical investigation of the OS-EM method. Moreover, we present some basic results concerning the KL-distance. Section 3 contains an analysis of the OS-EM iteration (1.10), i.e., monotonicity results and consequences concerning the asymptotic behavior of the iterations. Section 4 studies the case of noisy data and introduces the *loping OS-EM method* (4.4) which is a modification of the OS-EM iteration for noisy data. Stability results that use discrepancy type principles are stated. In section 5, we present some numerical experiments regarding application of the OS-EM methods to the inversion of the circular Radon transform. Section 6 is devoted to final remarks and conclusions.

2. Assumptions and basic results

Throughout this paper we assume the domains Ω and Σ in section 1 to be open bounded subsets of $\mathbb{R}^d$, $d \geqslant 1$. The parameter space for investigating system (1.6) is

$$\Delta := \left\{x \in L^1(\Omega);\, x \geqslant 0, \int_\Omega x(t)\,\mathrm{d}t = 1\right\}, \tag{2.1}$$

and the starting element x_0 of iteration (1.10) is chosen such that $x_0 \in \Delta$.

Moreover, we make the following assumptions to the framework introduced in section 1:

(A1) The kernel functions $a_j : \Sigma \times \Omega \to \mathbb{R}$, $j = 0, \ldots, N-1$, in (1.2) satisfy $\int_\Sigma a_j(s,t)\,\mathrm{d}s = 1$ for a.e. $t \in \Omega$.
(A2) There exist positive constants m and M such that $m \leqslant a_j(s,t) \leqslant M$ a.e. in $\Sigma \times \Omega$.
(A3) The exact data $y_j \in L^1(\Sigma)$ in (1.6) satisfy $\int_\Sigma y_j(s)\,\mathrm{d}s = 1$; moreover, there exists $M' > 0$ such that $y_j(s) \leqslant M'$ a.e. in Σ.
(A4) System (1.6) has a non-negative solution $x^* \in \Delta$, which does not vanish a.e. in Ω; moreover, $d(x^*, x_0) < \infty$.

Assumption (A2) implies that the operators $\mathcal{A}_j : L^1(\Omega) \to L^1(\Sigma)$ are continuous. In fact, when considered from $L^p(\Omega)$ into $L^p(\Sigma)$, they are also continuous. Moreover, any $\mathcal{A}_j x_k$ is in $L^\infty(\Sigma)$ and bounded away from zero. This further ensures that $1/\mathcal{A}_j x_k$ has the same properties and then yields that the integrals in (1.10) are well defined.

Remark 2.1. Assumption (A1) can be obtained by re-scaling of $\mathcal{A}_j$, provided that $(\mathcal{A}_j 1)(t)$ is independent of t. This independence, known as subset balance, is essential for the convergence proof in the finite-dimensional setting [1, 12]. Otherwise the iterations need not converge, even for exact data.

The boundedness away from zero assumption, which appears in (A2), was used also in [20–22] or, for a smoothing kernel, in [6]. Note that the bounds m and M are involved in the stopping rule analyzed in section 4. If no estimation of the bounds is available, then one can use a similar stopping rule which makes no use of the bounds—see remark 4.6. Moreover, if no positive lower bound exists at all, one can use a transformation of variables to obtain the boundedness away from zero—see remark 5.1 where a tomographic application is investigated.

In the following, we discuss some basic properties of the KL-distance in (1.9) that will be needed in the forthcoming sections. This functional plays a key role in the convergence analysis of the OS-EM method. For details, we refer the reader to [25, 26].

Lemma 2.2. *Let u and v be two L^1 functions such that (u, v) is in the domain of the KL-distance defined in (1.9). The following assertions hold true:*

(i) $d(v, u) \geqslant 0$ *and* $d(v, u) = 0$ *iff* $v = u$ *a.e.*
(ii) $\|v - u\|_{L^1}^2 \leqslant \left(\frac{2}{3}\|v\|_{L^1} + \frac{4}{3}\|u\|_{L^1}\right) d(v, u)$.
(iii) The function $(v, u) \mapsto d(v, u)$ *is convex.*
(iv) Let $\{v_n\}$ *and* $\{u_n\}$ *be given sequences in* L^1. *If* $\{u_n\}$ *is bounded and* $\lim_{n\to\infty} d(v_n, u_n) = 0$, *then* $\lim_{n\to\infty} \|v_n - u_n\|_{L^1} = 0$.

3. The OS-EM method for exact data

The first result of this section relates to a monotonicity property of the OS-EM iteration.

Lemma 3.1. *Let assumptions (A1)–(A3) be satisfied, and let $x \in \Delta$. Moreover, denote $P_j(x) = x\mathcal{A}_j^*(y_j/\mathcal{A}_j x)$. Then the following assertions hold true:*

(i) $P_j(x) \in \Delta$ *and* $d(P_j(x), x) \leqslant f_j(x) - f_j(P_j(x))$, *for* $j = 0, \ldots, N-1$;
(ii) If $x^* \in \Delta$ *is a minimizer of* f_j *for some* $0 \leqslant j \leqslant N-1$, *and* $d(x^*, x) < \infty$, *then* $d(x^*, P_j(x)) < \infty$ *and* $f_j(x) - f_j(x^*) \leqslant d(x^*, x) - d(x^*, P_j(x))$.

Proof. Results immediately from [26, proposition 3.1] applied to the function f_j and the corresponding P_j. □

From lemma 3.1 (i) and lemma 2.2 (i) we conclude that $f_j(P_j(x)) \leqslant f_j(x)$ and $P_j(x) \in \Delta$. Moreover, if $x^* \in \Delta$ is a solution to (1.6) with $d(x^*, x) < \infty$, then x^* minimizes f_j, for every $j = 0, \ldots, N-1$. Lemma 3.1 (ii) and the fact that $f_j(x^*) = 0$ therefore yield

$$f_j(x) \leqslant d(x^*, x) - d(x^*, P_j(x)), \qquad j = 0, \ldots, N-1. \tag{3.1}$$

In the next lemma, we reinterpret the inequalities derived in lemma 3.1 in terms of the OS-EM iteration.

Lemma 3.2. *Let assumptions (A1)–(A3) be satisfied and let $\{x_k\}$ be defined by iteration (1.10). Then the following assertions hold true:*

(i) $d(x_{k+1}, x_k) \leqslant f_{[k]}(x_k) - f_{[k]}(x_{k+1})$.
(ii) If $x^* \in \Delta$ *is a solution of (1.6), then* $f_{[k]}(x_k) \leqslant d(x^*, x_k) - d(x^*, x_{k+1})$.

Proof. Results from lemma 3.1 and (3.1). □

In the next theorem, we formulate the main monotonicity results for the OS-EM iteration with respect to the KL-distance, as well as convergence results in case the iterations are bounded.

Theorem 3.3. *Let assumptions (A1)–(A3) be satisfied, and the sequence $\{x_k\}$ be defined by iteration (1.10). Then we have*

(i) $f_{[k]}(x_{k+1}) \leqslant f_{[k]}(x_k)$*, for every* $k \in \mathbb{N}$.
Moreover, if assumption (A4) is satisfied, then the following assertions hold true:
(ii) The sequence $\{d(x^*, x_k)\}$ *is nonincreasing;*
(iii) $\lim_{k\to\infty} f_{[k]}(x_k) = 0$*;*
(iv) $\lim_{k\to\infty} d(x_{k+1}, x_k) = 0$*;*
(v) For each $0 \leqslant j \leqslant N-1$ *and* $p \in [1, \infty)$ *we have*

$$\lim_{m\to\infty} \|\mathcal{A}_j x_{j+mN} - y_j\|_{L^p(\Sigma)} = 0. \tag{3.2}$$

(vi) If $x_0 \in L^\infty(\Omega)$ *and* $\{x_k\}$ *is bounded in some* $L^p(\Omega)$ *space, with* $p \in (1, \infty)$*, then* $\{x_k\}$ *has a subsequence which converges weakly in* $L^p(\Omega)$ *to a solution of system (1.6).*

Proof. Item (i) follows from lemma 3.2 (i) and lemma 2.2 (i). Item (ii) follows from lemma 3.2 (ii), (1.8) and lemma 2.2 (i). Item (ii) implies the existence of $\mu \geqslant 0$ such that $\lim_{k\to\infty} d(x^*, x_k) = \mu$. Thus, (iii) follows from lemma 3.2 (ii).

To prove (iv), note that (i) and (iii) imply

$$\lim_{k\to\infty} f_{[k]}(x_{k+1}) = 0. \tag{3.3}$$

Now, (iv) results from (3.3), item (iii) and lemma 3.2 (i).

Next we prove (v). Since $f_j(x) = d(y_j, \mathcal{A}_j x)$, it follows from (iv) that $\lim_{k\to\infty} d(y_{[k]}, A_{[k]}x_k) = 0$. Consequently, $\lim_{m\to\infty} d(y_j, \mathcal{A}_j x_{j+mN}) = 0$, for every $j = 0, \ldots, N-1$. Now, by applying lemma 2.2 (iv) we obtain (3.2) for $p = 1$. The case $p \in (1, \infty)$ follows from [26, proposition 4.1 and lemma 4.2].

In order to prove assertion (vi), we first show the following:

Claim: $x_k \in L^\infty(\Omega)$ for each $k \in \mathbb{N}$.

Since $x_0 \in \Delta$ by hypothesis, we have $m \leqslant (\mathcal{A}_0 x_0)(s) \leqslant M$ a.e. in Σ by (A2). Consequently, it results from (A2) and (A3) that

$$\frac{a_1(s,t)y_0(s)}{(\mathcal{A}_0 x_0)(s)} \leqslant \frac{MM'}{m}, \qquad \text{a.e. in } \Sigma \times \Omega.$$

Hence, $x_1 \in L^\infty(\Omega)$ follows from $x_0 \in L^\infty(\Omega)$ and (1.10). Observe that $x_k \in L^\infty(\Omega)$ together with (A2) and (A3) imply $x_{k+1} \in L^\infty(\Omega)$. This, based on induction, completes the proof of the claim.

By hypothesis, the sequence $\{x_k\}$ is bounded in $L^p(\Omega)$, for some $p \in (1, +\infty)$. Therefore, there is a subsequence denoted by $\{x_l\}_{l\in\mathbb{N}}$, which converges weakly in $L^p(\Omega)$ to some $z \in L^p(\Omega)$. It remains to verify that z solves (1.6).

We distinguish two cases:

Case I: For each fixed $0 \leqslant j \leqslant N-1$, the subsequence $\{x_l\}_{l\in\mathbb{N}}$ obtained above contains infinitely many terms with indices of the form $l = j + mN, m \in \mathbb{N}$. Then, for $j = 0$, we can extract from $\{x_l\}$ a subsequence $\{x_{l_i}\}$ with indices of the form $l_i = m_i N$. Obviously, $[l_i] = 0$ for all indices of the subsequence $\{x_{l_i}\}$, and from (v) it follows that $\mathcal{A}_0 x_{l_i} \to y_0$ strongly as $i \to \infty$, and thus weakly in $L^p(\Sigma)$. Since $\mathcal{A}_0$ is continuous from $L^p(\Omega)$ to $L^p(\Sigma)$ due to (A2), it is weakly continuous. Therefore, $\mathcal{A}_0 x_{l_i} \to \mathcal{A}_0 z$ weakly in $L^p(\Sigma)$. From the uniqueness of weak limits it follows that $\mathcal{A}_0 z = y_0$.

By repeating the argumentation for $j = 1, \dots, N-1$, we conclude that $\mathcal{A}_j z = y_j$, for $j = 1, \dots, N-1$, thus proving that z is a solution to system (1.6).

Case II: There is at least one $0 \leqslant j_0 \leqslant N-1$ such that the subsequence $\{x_l\}_{l\in\mathbb{N}}$ obtained above contains infinitely many terms with indices of the form $l = j_0 + mN$, $m \in \mathbb{N}$. Arguing as in case I, we obtain a subsequence $\{x_{l_i}\}$, with indices of the form $l_i = j_0 + m_i N$, $i \in \mathbb{N}$, such that $\mathcal{A}_{j_0} x_{l_i} \to y_{j_0}$ weakly in $L^p(\Sigma)$, and conclude that $\mathcal{A}_{j_0} z = y_{j_0}$.

Now, let us consider the subsequence $\{x_{l_i+1}\}$, with indices of the form $l_i + 1 = (j_0+1) + m_i N$, $i \in \mathbb{N}$.[4] Being a subsequence of the weakly convergent sequence $\{x_l\}$, it also converges weakly to z in $L^p(\Omega)$. From the continuity of $\mathcal{A}_{[j_0+1]}$ follows $\mathcal{A}_{[j_0+1]} x_{l_i+1} \to \mathcal{A}_{[j_0+1]} z$ as $i \to \infty$, weakly in $L^p(\Sigma)$. Moreover, (v) yields $\mathcal{A}_{[j_0+1]} x_{l_i+1} \to y_{[j_0+1]}$ weakly in $L^p(\Sigma)$. Then we conclude that $\mathcal{A}_{[j_0+1]} z = y_{[j_0+1]}$.

Repeating the argumentation for the subsequences $\{x_{l_i+2}\}, \dots, \{x_{l_i+N-1}\}$, we conclude that $\mathcal{A}_j z = y_j$, for every j, thus proving that z is a solution to system (1.6). □

Remark 3.4. We can interpret theorem 3.3 (v) as follows: if we consider the subsequence $\{x_{j+mN}\}_{m\in\mathbb{N}}$ formed by the jth component of each cycle of the OS-EM iteration (where $0 \leqslant j \leqslant N-1$), then the L^1-norm of the residual corresponding to this subsequence converges to zero.

Moreover, theorem 3.3 (iv) guarantees that, given any two 'consecutive' subsequences $\{x_{j+mN}\}_{m\in\mathbb{N}}$ and $\{x_{(j+1)+mN}\}_{m\in\mathbb{N}}$, we have

$$\lim_{m\to\infty} d(x_{(j+1)+mN}, x_{j+mN}) = 0,$$

for each $j = 0, \dots, N-2$.

In the finite-dimensional case, the boundedness assumption (vi) is not needed, as monotonicity of the distances $d(x^*, x_k)$ guarantees it (see, e.g., [13]).

4. The loping OS-EM method for noisy data

Our next goal is to modify the OS-EM iteration by introducing a relaxation parameter, and to investigate monotonicity and stability results for this modified iteration (the so-called *loping OS-EM method*) in the case of noisy data. As remarked in [12], 'With noisy data though, inconsistent applications (of discrete OS-EM—authors' note) result.'

We aim at characterizing the loping OS-EM method as an *iterative regularization method* in the sense of [7].

For the rest of this section we assume that the right-hand side of (1.6) is not exactly known. Instead, we have only approximate measured data $y_j^\delta \in L^1(\Sigma)$ satisfying

$$\left\| y_j - y_j^\delta \right\|_{L^1} \leqslant \delta_j, \qquad j = 0, \dots, N-1. \tag{4.1}$$

We denote $\delta := (\delta_0, \dots, \delta_{N-1})$.

In this noisy data case we are interested in finding an approximate solution for the system

$$\mathcal{A}_j x = y_j^\delta, \qquad j = 0, \dots, N-1. \tag{4.2}$$

The following assumptions are required for the analysis:

(A5) The noisy data $y_j^\delta \in L^1(\Sigma)$ satisfies $\int_\Sigma y_j^\delta(s)\, \mathrm{d}s = 1$.

(A6) There exist $M_1, m_1 > 0$ such that $M_1 \geqslant y_j^\delta \geqslant m_1$ a.e. in Σ.

[4] Note that $\{x_{l_i+1}\}$ may contain infinitely many elements which do not belong to the convergent subsequence $\{x_l\}$, while $\{x_{l_i}\}$ is a subsequence extracted from $\{x_l\}$.

It is worth mentioning that (A5) is stated only for simplicity of the presentation and is not a restrictive assumption, since the noisy data can always be normalized. Due to assumption (A3), the new noise level increases at most by a factor 2.

Also necessary for the analysis are the following functions associated with the equations of system (4.2)

$$f_j^\delta(x) := \int_\Sigma \left[y_j^\delta(s) \ln \frac{y_j^\delta(s)}{(\mathcal{A}_j x)(s)} - y_j^\delta(s) + (\mathcal{A}_j x)(s) \right] \mathrm{d}s. \tag{4.3}$$

Note that $f_j^\delta(x) = d\big(y_j^\delta, \mathcal{A}_j x\big)$.

The *loping OS-EM iteration* for the inverse problem (4.2) with noisy data is defined by

$$x_{k+1}^\delta = x_k^\delta \omega_k \tag{4.4a}$$

where

$$\omega_k = \begin{cases} \displaystyle\int_\Sigma \frac{a_{[k]}(s,\cdot) y_{[k]}^\delta(\cdot)}{\big(\mathcal{A}_{[k]} x_k^\delta\big)(s)} \,\mathrm{d}s =: P_{[k]}^\delta\big(x_k^\delta\big), & f_{[k]}^\delta\big(x_k^\delta\big) > \tau\gamma\delta_{[k]} \\ 1, & \text{else}. \end{cases} \tag{4.4b}$$

The constants τ and γ in (4.4*b*) are chosen such that

$$\tau > 1, \qquad \gamma = \max\left\{ \left| \ln \frac{m_1}{M} \right|, \left| \ln \frac{M_1}{m} \right| \right\}, \tag{4.5}$$

where m, M, m_1, M_1 are the positive constants defined in (A2) and (A6).

The term 'loping' reflects the fact that the kth iterative update in (4.4) is loped (omitted), if the corresponding residual $f_{[k]}^\delta\big(x_k^\delta\big)$ is below a certain threshold.

Remark 4.1. It is worth noting that, for noisy data, the iteration in (4.4) is much different from the iteration in (1.10): the relaxation parameter ω_k effects that the iterates defined in (4.4*a*) become stationary if all components of the residual vector $d\big(y_{[k]}^\delta, \mathcal{A}_{[k]} x_k^\delta\big)$ fall below a pre-specified threshold.

Another consequence of using these relaxation parameters is the fact that, after a large number of iterations, $\omega_k = 1$ for some k within each iteration cycle. Therefore, the computational evaluation of the adjoint operator

$$\mathcal{A}_{[k]}^* \left(\frac{y_{[k]}^\delta}{\mathcal{A}_{[k]} x_k^\delta} \right) = \int_\Sigma \frac{a_{[k]}(s,\cdot) y_{[k]}^\delta(\cdot)}{\big(\mathcal{A}_{[k]} x_k^\delta\big)(s)} \,\mathrm{d}s$$

might be loped, making the loping OS-EM iteration in (4.4) a fast alternative to the OS-EM method.

In the case of noise free data, i.e. $\delta_j = 0$ in (4.1), we choose $\omega_k = P_{[k]}^\delta\big(x_k^\delta\big) = P_{[k]}(x_k)$ and the loping OS-EM iteration (4.4) reduces to the OS-EM method (1.10).

In the following, we prove a monotonicity result for the loping OS-EM iteration in the case of noisy data. First however, we derive an auxiliary estimate.

Lemma 4.2. *Let assumptions (A1)–(A5) hold true. Moreover, let y_j^δ, δ_j be given as in (4.1), with $\delta_{j_0} > 0$ for some $0 \leqslant j_0 \leqslant N-1$. Then we have*

$$f_{[k]}^\delta\big(x_k^\delta\big) - d\big(y_{[k]}, y_{[k]}^\delta\big) \leqslant d\big(x^*, x_k^\delta\big) - d\big(x^*, x_{k+1}^\delta\big). \tag{4.6}$$

for all $k \in \mathbb{N}$ with $[k] = j_0$.

Proof. Since (A1)–(A3) are satisfied, we argue as in the proof of [26, proposition 5.2] to conclude that for every $v, w \in \Delta$, and $0 \leqslant j \leqslant N-1$ the inequality

$$d(P_j(w), P_j^\delta(v)) \leqslant d(y_j, y_j^\delta) + d(P_j(w), v) - d(P_j(w), w) + f_j(w) - f_j(v),$$

holds true. Therefore, given $k \in \mathbb{N}$ with $[k] = j_0$, (4.6) follows by taking $j = [k]$, $w = x^*$, $v = x_k^\delta$, and by observing that $P_{[k]}(x^*) = x^*$. □

Proposition 4.3. *Let assumptions (A1)–(A6) hold true and τ, γ be defined as in (4.5). Moreover, let y_j^δ, δ_j be given as in (4.1) with $\delta_j > 0$ for $j = 0, \dots, N-1$. Then the sequence $\{x_k^\delta\}$ defined by iteration (4.4) satisfies*

$$d\left(x^*, x_{k+1}^\delta\right) \leqslant d\left(x^*, x_k^\delta\right), \qquad k \in \mathbb{N}. \tag{4.7}$$

Proof. If $f_{[k]}^\delta\left(x_k^\delta\right) \leqslant \tau\gamma\delta_{[k]}$, then $w_k = 1$ by (4.4*b*). Therefore, $x_{k+1}^\delta = x_k^\delta$ and (4.7) follows with equality. If $f_{[k]}^\delta\left(x_k^\delta\right) > \tau\gamma\delta_{[k]}$, note that a simple calculation yields

$$d\left(x^*, x_k^\delta\right) - d\left(x^*, x_{k+1}^\delta\right) \geqslant f_{[k]}^\delta\left(x_k^\delta\right) + \int_\Sigma [y_{[k]} - y_{[k]}^\delta] \ln\left(\frac{y_{[k]}^\delta}{(\mathcal{A}_{[k]} x_k^\delta)}\right) \mathrm{d}s$$

from (4.6). Therefore, (4.7) follows from

$$\begin{aligned} f_{[k]}^\delta\left(x_k^\delta\right) + \int_\Sigma \left[y_{[k]} - y_{[k]}^\delta\right] \ln\left(\frac{y_{[k]}^\delta}{\mathcal{A}_{[k]} x_k^\delta}\right) \mathrm{d}s &\geqslant f_{[k]}^\delta\left(x_k^\delta\right) - \left\|y_{[k]} - y_{[k]}^\delta\right\|_{L^1} \left\| \ln\left(\frac{y_{[k]}^\delta}{\mathcal{A}_{[k]} x_k^\delta}\right)\right\|_{L^\infty} \\ &\geqslant f_{[k]}^\delta\left(x_k^\delta\right) - \delta_{[k]} \max\left\{\left|\ln\frac{m_1}{M}\right|, \left|\ln\frac{M_1}{m}\right|\right\} \\ &\geqslant f_{[k]}^\delta\left(x_k^\delta\right) - \gamma\delta_{[k]} \\ &\geqslant (\tau - 1)\gamma\delta_{[k]} \end{aligned} \tag{4.8}$$

together with (4.5). To obtain the inequalities above we used (4.1), (4.5), (A2) and (A6). □

Proposition 4.3 gives us a hint on how to choose the stopping rule for the loping OS-EM iteration. That is, we stop the iteration at

$$k_*^\delta := \min\left\{mN \in \mathbb{N};\, x_{mN}^\delta = x_{mN+1}^\delta = \cdots = x_{mN+N}^\delta\right\}. \tag{4.9}$$

In other words, k_*^δ is the smallest integer multiple of N such that

$$x_{k_*^\delta} = x_{k_*^\delta+1} = \cdots = x_{k_*^\delta+N}. \tag{4.10}$$

In the following, we prove that the stopping index k_*^δ in (4.9) is well defined and that the corresponding iterations stably converge to a solution of the system, if they are bounded in some L^p space with $p \in (1, +\infty)$.

Theorem 4.4. *Let assumptions (A1)–(A6) be satisfied, and $k_*^\delta \in \mathbb{N}$ be chosen according to (4.9). Then the following assertions hold true:*

(i) The stopping index k_*^δ defined in (4.9) is finite.
(ii) More precisely, $k_*^\delta = O\left(\delta_{\min}^{-1}\right)$, where $\delta_{\min} := \min\{\delta_0, \dots, \delta_{N-1}\}$.
(iii) $d\left(y_j^\delta, \mathcal{A}_j x_{k_*^\delta}^\delta\right) \leqslant \tau\gamma\delta_j$, for every $j = 0, \dots, N-1$.
(iv) For every $p \in [1, +\infty)$ and every $j = 0, \dots, N-1$ we have

$$\lim_{\delta \to 0} \left\|\mathcal{A}_j x_{k_*^\delta}^\delta - y_j\right\|_{L^p(\Sigma)} = 0.$$

(v) Let $\{\delta^l := (\delta_0^l, \ldots, \delta_{N-1}^l)\}_{l\in\mathbb{N}}$ *be a sequence in* $(0,\infty)^N$ *with* $\lim_{l\to\infty} \delta_j^l = 0$*, for each* $0 \leqslant j \leqslant N-1$*. Moreover, let* $\{y^l := (y_0^l, \ldots, y_{N-1}^l)\}_{l\in\mathbb{N}}$ *be a sequence of noisy data satisfying*

$$\left\| y_j - y_j^l \right\|_{L^1} \leqslant \delta_j^l, \qquad j = 0, \ldots, N-1, \qquad l \in \mathbb{N}, \tag{4.11}$$

and $k_*^l := k_*^{\delta^l} = k_*(\delta^l, y^l)$ *denote the corresponding stopping index defined in (4.9). If* $x_0 \in L^\infty(\Omega)$ *and the sequence* $\{x_{k_*^l}^{\delta^l}\}_{l\in\mathbb{N}}$ *is bounded in some* $L^p(\Omega)$ *space, with* $p \in (1, +\infty)$*, then it has a subsequence which converges weakly in* $L^p(\Omega)$ *to a solution to system (1.6).*

Proof. (i) Assume by contradiction that k_*^δ is not finite. Then it results from (4.9) that $x_{k+1}^\delta \neq x_k^\delta$ at least once in each cycle of iteration (4.4). Hence for every $m \in \mathbb{N}$ there exists $j_m \in \{0, \ldots, N-1\}$ such that

$$f_{j_m}^\delta\left(x_{j_m+mN}^\delta\right) > \tau\gamma\delta_{j_m}. \tag{4.12}$$

From (4.8) in the proof of proposition 4.3, it follows

$$d\left(x^*, x_k^\delta\right) - d\left(x^*, x_{k+1}^\delta\right) \geqslant \max\left\{f_{[k]}^\delta\left(x_k^\delta\right) - \gamma\delta_{[k]}, 0\right\}, \qquad k \in \mathbb{N}.$$

Summing up this inequality for $k = 0, \ldots, lN-1$ implies[5]

$$\begin{aligned} d(x^*, x_0) - d\left(x^*, x_{lN}^\delta\right) &\geqslant \sum_{k=0}^{lN-1} \max\left\{f_{[k]}^\delta\left(x_k^\delta\right) - \gamma\delta_{[k]}, 0\right\} \\ &= \sum_{m=0}^{l}\sum_{j=0}^{N-1} \max\left\{f_j^\delta\left(x_{j+mN}^\delta\right) - \gamma\delta_j, 0\right\}, \qquad l \in \mathbb{N}. \end{aligned}$$

Then, it follows from (4.12)

$$d(x^*, x_0) \geqslant \sum_{m=0}^{l}\left(f_{j_m}^\delta\left(x_{j_m+mN}^\delta\right) - \gamma\delta_{j_m}\right) > \sum_{m=0}^{l}(\tau-1)\gamma\delta_{j_m} > l(\tau-1)\gamma\delta_{\min}, \qquad l \in \mathbb{N}. \tag{4.13}$$

However, due to (4.5), the right-hand side of (4.13) becomes unbounded as $l \to \infty$, contradicting (A4). Therefore, k_*^δ must be finite. To prove (ii), it is enough to take $l = k_*^\delta/N \in \mathbb{N}$ in (4.13) and obtain $k_*^\delta < Nd(x^*, x_0)/((\tau-1)\gamma\delta_{\min})$.

To prove (iii), we assume by contradiction that

$$f_{j_0}^\delta\left(x_{k_*^\delta}^\delta\right) = d\left(y_{j_0}^\delta, \mathcal{A}_{j_0} x_{k_*^\delta}^\delta\right) > \tau\gamma\delta_{j_0},$$

for some $0 \leqslant j_0 \leqslant N-1$. Thus, it results from (4.10) that $f_{j_0}^\delta\left(x_{k_*^\delta+j_0}^\delta\right) > \tau\gamma\delta_{j_0}$. Therefore, it follows from (4.8) in the proof of proposition 4.3 that

$$0 = d\left(x^*, x_{k_*^\delta+j_0}^\delta\right) - d\left(x^*, x_{k_*^\delta+j_0+1}^\delta\right) \geqslant f_{j_0}^\delta\left(x_{k_*^\delta+j_0}^\delta\right) - \gamma\delta_{j_0} \geqslant (\tau-1)\gamma\delta_{j_0},$$

which contradicts (4.5).

(iv) and (v) The proofs follow the lines of the proof of theorem 3.3 (v), (vi). □

As pointed out in remark 3.4, in finite-dimensional spaces, one does not need to assume boundedness of the iterates (part of the hypothesis (v)), since this is a consequence of the intermediary monotonicity results.

[5] Note that $x_0^\delta = x_0$.

Remark 4.5 (Stability for noisy data in L^2)**.** When dealing with inverse problems, bounds for the noisy data are most commonly given in the L^2-norm, i.e. the approximate measured data $y_j^\delta \in L^2(\Sigma)$ are assumed to satisfy

$$\|y_j - y_j^\delta\|_{L^2} \leqslant \delta_j, \qquad j = 0, \ldots, N-1, \tag{4.14}$$

instead of (4.1). In this case, the loping OS-EM iteration is defined by (4.4), where the 'loping condition' $f_{[k]}^\delta(x_k^\delta) > \tau\gamma\delta_{[k]}$ in (4.4*b*) is substituted by

$$f_{[k]}^\delta(x_k^\delta) > \tau\delta_{[k]}\gamma_k, \qquad \gamma_k := \left\|\ln\left(y_{[k]}^\delta / \left(\mathcal{A}_{[k]}x_k^\delta\right)\right)\right\|_{L^2}. \tag{4.15}$$

Under these assumptions it is possible to state a stability result, similar to that in theorem 4.4 (iv). One argues as follows:

- First of all, note that monotonicity of the error with respect to the KL-distance (as in (4.7)) follows when using the Cauchy–Schwarz inequality in $L^2(\Sigma)$ to derive the estimate (compare with (4.8))

$$f_{[k]}^\delta(x_k^\delta) + \int_\Sigma \left[y_{[k]} - y_{[k]}^\delta\right] \ln\left(y_{[k]}^\delta / \left(\mathcal{A}_{[k]}x_k^\delta\right)\right) \mathrm{d}s \geqslant (\tau - 1)\delta_{[k]}\gamma_k.$$

- By defining the stopping index k_*^δ as in (4.9), its finiteness can be proven analogously as in theorem 4.4 (i)—provided that γ_k can be bounded from below by some positive constant. Moreover, the following estimate holds true (compare with item (iii) of theorem 4.4)

$$d\left(y_j^\delta, \mathcal{A}_j x_{k_*^\delta}^\delta\right) \leqslant \tau\delta_j \left\|\ln\left(y_{[k]}^\delta / \left(\mathcal{A}_{[k]}x_{k_*^\delta}^\delta\right)\right)\right\|_{L^2}, \qquad j = 0, \ldots, N-1. \tag{4.16}$$

- In theorem 4.4 (iv), if one substitutes the assumption (4.11) by $\|y_j - y_j^l\|_{L^2} \leqslant \delta_j^l$, $j = 0, \ldots, N-1, l \in \mathbb{N}$, then the proof of the stability result carries on with analogous argumentation.

Note that the estimate in (4.16) allows for the following interpretation: the loping OS-EM iteration should be stopped at the index k_*^δ (an integer multiple of N) when for the first time (4.16) is satisfied within a whole cycle.

Remark 4.6. In the case that the error estimates for the noisy data are given in an L^q norm with $q \in [1, +\infty)$, one can work with a stopping rule similar to (4.15) depending on parameters $\tilde{\gamma}_k := \left\|\ln\left(y_{[k]}^\delta / \left(\mathcal{A}_{[k]}x_k^\delta\right)\right)\right\|_{L^p}$, where $p \in (1, +\infty]$ is such that $p^{-1} + q^{-1} = 1$. The advantage of using this type of stopping rule resides on the fact that no quantitative information on the constants m, M, m_1, M_1 is required to compute the iteration. In other words, the constant γ is not required neither to test the 'loping condition' (4.15) nor to verify the stopping rule based on (4.16). This is obviously not the case if the 'loping condition' $f_{[k]}^\delta(x_k^\delta) > \tau\gamma\delta_{[k]}$ with γ defined by (4.5) is to be implemented.

5. Numerical example

In this section, we compare the numerical performance of our loping OS-EM method with the OS-EM and EM methods. As a benchmark problem we use a system of linear equations for the circular Radon transform. The inversion of the circular Radon is relevant for the emerging *photoacoustic computed tomography* [17, 23, 28, 30].

Let $\epsilon < 1$ be some small positive number, let $\Omega := B_{1-\epsilon}(0) \subset \mathbb{R}^2$ denote the disc with radius $1 - \epsilon$ centered at the origin, set

$$\Sigma_j := \left(\frac{2j\pi}{N}, \frac{2(j+1)\pi}{N}\right) \times (0, 2), \qquad j = 0, \ldots, N-1,$$

and let $\Phi : \mathbb{R} \to \mathbb{R}$ be a continuous nonnegative function with $\operatorname{supp}(\Phi) = [-\epsilon, \epsilon]$ and $\int_{\mathbb{R}} \Phi = 1$.

Our aim is the stable solution to (1.6), with $\mathcal{A}_j x := \Phi *_r (\mathcal{M}_j x)$, where

$$(\mathcal{M}_j x)(\varphi, r) := \frac{rN}{2\pi} \int_{S^1} x((\cos\varphi, \sin\varphi) + r\omega)\, \mathrm{d}\omega, \qquad (\varphi, r) \in \Sigma_j \tag{5.1}$$

is the *circular Radon transform* restricted to Σ_j, S^1 is the unit sphere in $\mathbb{R}^2$ and $\Phi *_r y = \mathcal{I}_\Phi y$ denotes the convolution of Φ and y . In (5.1), x is considered as an element in $L^1(\mathbb{R}^2)$ by extending it with zero outside of Ω.

One verifies that the operators $\mathcal{A}_j$ can be written in the form (1.7), with $s = (\varphi, r)$ and

$$a_j(\varphi, r, t) = \Phi\left(|(\cos\varphi, \sin\varphi) - t| - r\right), \qquad j = 0, \ldots, N-1.$$

Moreover, the adjoint of $\mathcal{A}_j$ is given by $\mathcal{A}_j^* y = \mathcal{B}_j(\Phi *_r y)$, where

$$(\mathcal{B}_j y)(t) = \frac{N}{2\pi} \int_{2j\pi/N}^{2(j+1)\pi/N} y(|t - (\cos\varphi, \sin\varphi)|)\, \mathrm{d}\varphi \tag{5.2}$$

is the *circular backprojection*. Hence $\mathcal{A}_j^* 1 = 1$ and the operators $\mathcal{A}_j$ satisfy assumption (A1). However, since a_j are not bounded from below, $\mathcal{A}_j$ do not satisfy (A2), but they can be appropriately modified as follows.

Remark 5.1. For any positive λ, the operators

$$\mathcal{A}_j^{(\lambda)} x := \frac{1}{1 + \lambda|\Sigma_j|} \left(\mathcal{A}_j x + \lambda \int_\Omega x \right), \qquad j = 0, \ldots, N-1,$$

clearly satisfy (A2). Since $(\mathcal{A}_j^{(\lambda)})^* y = \left(\mathcal{A}_j^* y + \lambda \int_{\Sigma_j} y\right)/(1 + \lambda|\Sigma_j|)$ we have $\left(\mathcal{A}_j^{(\lambda)}\right)^* 1 = 1$, proving that (A1) is also satisfied.

For the reasons shown in the above remark, we shall consider for the rest of this section the system of equations

$$\mathcal{A}_j^{(\lambda)} x = y_j^{(\lambda)} := \frac{1}{1 + \lambda|\Sigma_j|} \left(y_j + \lambda \int_{\Sigma_j} y_j \right), \qquad j = 0, \ldots, N-1. \tag{5.3}$$

The identity $\int_\Omega x = \int_\Omega x \mathcal{A}_j^* 1 = \int_{\Sigma_j} \mathcal{A}_j x$ implies that x is a solution to (5.3) if and only if x satisfies $\mathcal{A}_j x = y_j$.

If noisy data y_j^δ with $\|y_j^\delta - y_j\|_{L^1} \leqslant \delta_j$ are available, then

$$\left\|y_j^{(\lambda),\delta} - y_j^{(\lambda)}\right\|_{L^1} \leqslant \frac{1}{1 + \lambda|\Sigma_j|} \left(\left\|y_j^\delta - y_j\right\|_{L^1} + \lambda \left| \int_{\Sigma_j} y_j^\delta - y_j \right| \right) \leqslant \frac{\delta_j(1 + \lambda)}{1 + \lambda|\Sigma_j|},$$

where $y_j^{(\lambda),\delta}$ is defined in the same way as $y_j^{(\lambda)}$, with y_j replaced by y_j^δ. Therefore, the loping OS-EM iteration with noisy data $y_j^{(\lambda),\delta}$ applied to system (5.3) reads as

$$x_{k+1}^\delta := x_k^\delta \omega_k,$$
$$\omega_k := \begin{cases} \dfrac{\mathcal{B}_{[k]} \mathcal{I}_\Phi + \lambda \|\cdot\|_{L^1}}{1 + \lambda|\Sigma_j|} \left(\dfrac{y_{[k]}^\delta + \lambda}{\mathcal{I}_\Phi \mathcal{M}_{[k]} x_k^\delta + \lambda} \right), & \dfrac{d\left(y_{[k]}^\delta + \lambda, \mathcal{I}_\Phi \mathcal{M}_{[k]} x_k^\delta + \lambda\right)}{1 + \lambda} > \tau\gamma\delta_{[k]}, \\ 1, & \text{else}. \end{cases} \tag{5.4}$$

Here we made use of the fact that the initial guess satisfies $\int_\Omega x_0^\delta = 1$, which implies $\int_\Omega x_k^\delta = \int_{\Sigma_{[k]}} \mathcal{A}_{[k]} x_k^\delta = 1$ for every k.

Since the operators $\mathcal{A}_j^{(\lambda)}$ and the data $y_j^{(\lambda)}$, $y_j^{(\lambda),\delta}$ satisfy assumptions (A1)–(A6), the theory of the previous sections (we refer mainly to theorems 3.3 and 4.4) applies to the corresponding modified iteration (5.4).

Remark 5.2. Iteration (5.4) assumes continuous data $y_j^\delta \in L^1(\Sigma_j)$, whereas in practical applications only discrete data are available. In the following, we assume that data

$$\boldsymbol{y}_j^\delta[i_\varphi, i_r] := y_j^\delta(\boldsymbol{\varphi}[i_\varphi], \boldsymbol{r}[i_r]), \qquad (i_\varphi, i_r) \in \{jN_\varphi, \dots, (j+1)N_\varphi - 1\} \times \{0, \dots, N_r\},$$

are given, with $\boldsymbol{\varphi}[i_\varphi] := 2i_\varphi\pi/(N_\varphi N)$, $\boldsymbol{r}[i_r] := 2i_r/N_r$, and N_φ, $N_r + 1$ denoting the number of samples of y_j^δ in the angular and radial variables, respectively.

In the numerical implementation $\mathcal{M}_j, \mathcal{B}_j, \mathcal{I}_\Phi$ and d are replaced (as described below) with finite-dimensional approximations $\mathbf{M}_j, \mathbf{B}_j, \mathbf{I}_\Phi, \boldsymbol{d}$, and (5.4) is approximated by

$$x_{k+1}^\delta(\boldsymbol{t}[i]) \simeq \boldsymbol{x}_{k+1}^\delta[i] := \boldsymbol{x}_k^\delta[i]\boldsymbol{\omega}_k[i], \qquad i \in \{0, \dots, N_t\}^2$$

$$\boldsymbol{\omega}_k := \begin{cases} \dfrac{\mathbf{B}_{[k]}\mathbf{I}_\Phi + 4\pi\lambda/(N_r N_\varphi)\sum_{i_\varphi, i_r}(\cdot)}{1 + 4\pi\lambda/(N_r N_\varphi)} \left(\dfrac{\boldsymbol{y}_{[k]}^\delta + \lambda}{\mathbf{I}_\Phi \mathbf{M}_{[k]}\boldsymbol{x}_k^\delta + \lambda}\right), & \dfrac{\boldsymbol{d}\big(\boldsymbol{y}_{[k]}^\delta + \lambda, \mathbf{I}_\Phi \mathbf{M}_{[k]}^{(\lambda)}\boldsymbol{x}_k^\delta + \lambda\big)}{1 + \lambda} > \tau\gamma\delta_{[k]}^{(\lambda)}, \\ 1, & \text{else.} \end{cases} \tag{5.5}$$

Here $\boldsymbol{y}_j^\delta = \big(\boldsymbol{y}_j^\delta[i_\varphi, i_r]\big)_{i_\varphi, i_r}$, $\boldsymbol{x}_k^\delta = \big(\boldsymbol{x}_k^\delta[i]\big)_i$, $\boldsymbol{\omega}_k = (\boldsymbol{\omega}_k[i])_i$, and $\boldsymbol{t}[i] = -(1, 1) + 2i/N_t$ with $(N_t + 1)^2$ denoting the number of samples in the variable t.

(1) The discretized circular Radon transform

$$\mathbf{M}_j : \mathbb{R}^{(N_t+1)\times(N_t+1)} \to \mathbb{R}^{N_\varphi \times (N_r+1)}$$

is obtained by replacing x in (5.1) with the bilinear spline $T(\boldsymbol{x})$ satisfying $T(\boldsymbol{x})(\boldsymbol{t}[i]) = \boldsymbol{x}[i]$, and approximating the resulting integrals over the S^1 with the trapezoidal rule. This leads to

$$(\mathbf{M}_j\boldsymbol{x})[i_\varphi, i_r] = \frac{2}{N_t}\sum_{i_\omega=0}^{3\boldsymbol{r}[i_r]N_t} T(\boldsymbol{x})(\boldsymbol{\sigma}[i_\varphi] + \boldsymbol{r}[i_r]\boldsymbol{\omega}[i_\omega]), \tag{5.6}$$

where $\boldsymbol{\sigma}[i_\varphi] := (\cos\boldsymbol{\varphi}[i_\varphi], \sin\boldsymbol{\varphi}[i_\varphi])$, $\boldsymbol{\omega}[i_\omega] := (\cos(2\pi i_\omega/N_t), \sin(2\pi i_\omega/N_t))$, and $3\boldsymbol{r}[i_r]N_t$ is the number of supporting points when applying the trapezoidal rule.

(2) Assuming that $\epsilon = 2K/N_r$ for some $K \in \mathbb{N}$, the convolution $\mathcal{I}_\Phi y = \Phi *_r y$ is approximated by

$$(\mathbf{I}_\Phi\boldsymbol{y})[i_\varphi, i_r] = \frac{2}{N_r}\sum_{i_r'=i_r-K}^{i_r+K}\Phi\big(2(i_r' - i_r)/N_r\big)\boldsymbol{y}[i_\varphi, i_r'],$$

where $\boldsymbol{y}[i_\varphi, i_r] := 0$ for i_r outside $\{0, \dots, N_r\}$.

(3) The discretized back-projection $\mathbf{B}_j : \mathbb{R}^{N_\varphi\times(N_r+1)} \to \mathbb{R}^{(N_t+1)\times(N_t+1)}$ is defined by

$$(\mathbf{B}_j\boldsymbol{y})[i] := \frac{1}{N_\varphi}\sum_{i_\varphi=jN_\varphi}^{(j+1)N_\varphi-1} T_r(\boldsymbol{y})\big(i_\varphi, |\boldsymbol{t}[i] - \boldsymbol{\sigma}[i_\varphi]|\big),$$

if $\boldsymbol{t}[i] \in \Omega$, and setting $(\mathbf{B}_j\boldsymbol{y})[i] := 0$ for $\boldsymbol{t}[i] \notin \Omega$. Here $T_r(\boldsymbol{y})$ denotes the piecewise linear spline in the second variable satisfying $T_r(\boldsymbol{y})\big(i_\varphi, \boldsymbol{r}[i_r]\big) = \boldsymbol{y}[i_\varphi, i_r]$.

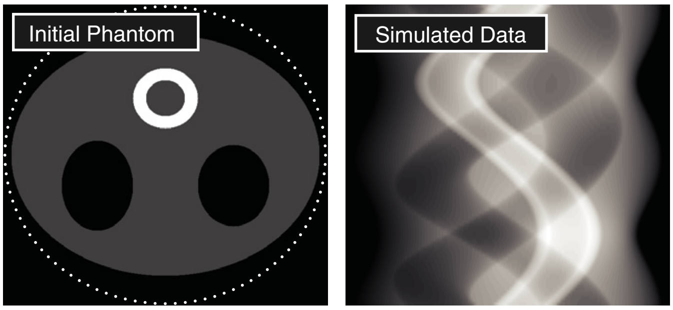

Figure 1. Original phantom x^* (left) and simulated data $(\mathcal{A}_j x^*)_j$.

(4) Finally, the discrete KL-distance is defined by

$$\boldsymbol{d}(\boldsymbol{v}, \boldsymbol{u}) = \frac{4\pi}{N_r N_\varphi N} \sum_{i_\varphi = jN_\varphi}^{(j+1)N_\varphi - 1} \sum_{i_r=0}^{N_r} \boldsymbol{v}[i_\varphi, i_r] \ln \frac{\boldsymbol{v}[i_\varphi, i_r]}{\boldsymbol{u}[i_\varphi, i_r]} - \boldsymbol{v}[i_\varphi, i_r] + \boldsymbol{u}[i_\varphi, i_r]$$

for $\boldsymbol{v}, \boldsymbol{u} \in \mathbb{R}^{N_\varphi \times (N_r+1)}$.

Remark 5.3. (Numerical complexity) Assuming $N_t = N_r$, the numerical complexity for performing one iteration cycle (which consists of N subsequent steps in (5.5)) is $\mathcal{O}\big(N_{\text{angle}} N_t^2\big)$. Here $N_{\text{angle}} = N_\varphi N$ corresponds to the overall angular data samples, which is independent of N in practice. Therefore, in the following we always compare the reconstruction error in dependence of the number of iteration cycles.

In the following numerical examples we apply the (loping) OS-EM iteration with $N = 1$ (corresponding to the EM algorithm), $N = 5$, $N = 10$ and $N = 20$ subsets. The original phantom x^* (the exact nonnegative solution) is shown in the left picture in figure 1 and consists of a superposition of characteristic functions. Note that a similar phantom was reconstructed in [12] where the OS-EM technique was introduced. The data $\boldsymbol{y}_j$, shown in the right picture in figure 1, were calculated numerically by (5.6) for $N_{\text{angle}} = N_\varphi N = 100$ angular samples. In order to avoid inverse crimes, much larger N_t is used for the data simulation as for the application of the loping OS-EM iteration. In all examples $\boldsymbol{x}_0 = \boldsymbol{x}_0^\delta = 1/((1-\epsilon)^2\pi)$ is used as initial guess and the parameters ϵ and λ are chosen to be 0.02 and 0.01, respectively.

The iterations of the OS-EM method applied to exact data with $N_t = N_r = 100$ and different values of N are depicted in figure 2. It can be seen that the fifth iterate with EM has similar quality as the first iterate with OS-EM for $N = 5$. As a rough rule one can say that making N cycles with the EM algorithm leads to an improvement similar to 1 cycle with the OS-EM algorithm. This can also be recognized in the left image in figure 3, where the evolution of the error is depicted with respect to the KL-distance.

In order to investigate the dependence of the OS-EM iteration on the discretization level, we repeated the experiment with $N_t = N_r = 200$. The right image in figure 3 shows the corresponding logarithmic error. As expected, the error is relatively independent on the discretization.

In the case of noisy data we apply the loping OS-EM iteration (5.5). The noisy data $\boldsymbol{y}_j^\delta$ are created by adding 5% Poisson distributed noise to the simulated data $\boldsymbol{y}_j$, such that $4\pi/(N_r N_\varphi) \sum \big|\boldsymbol{y}_j[i_\varphi, i_r] - \boldsymbol{y}_j^\delta[i_\varphi, i_r]\big| \cong 0.05$.

Remark 5.4. Our numerical experiments show that, for large δ and $\tau \cong 1$, far too many iterations are loped. A significant improvement can be obtained if $\tau = \tau(\delta)$ is chosen in

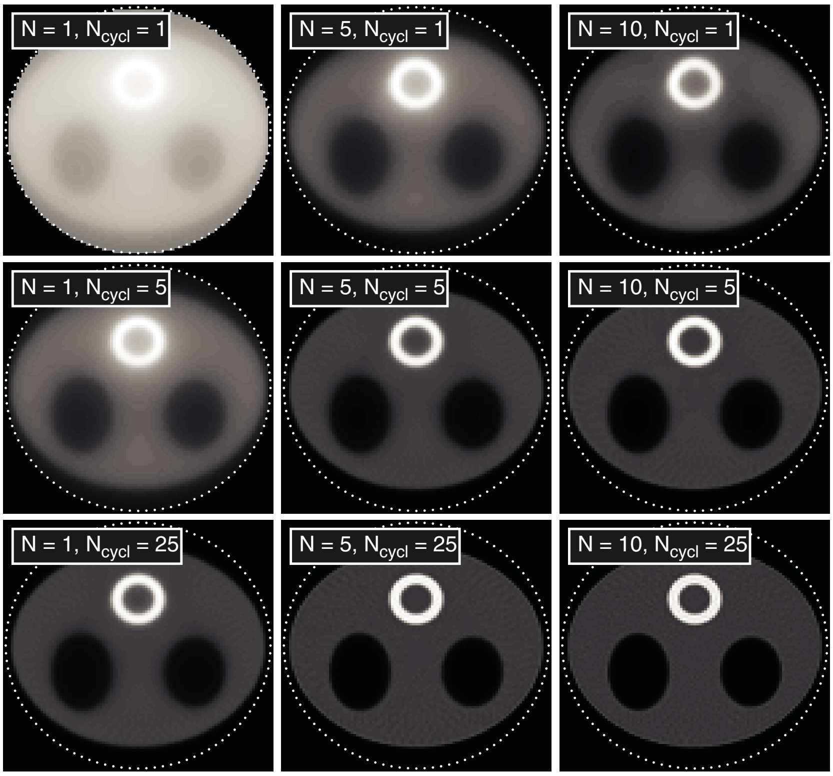

Figure 2. Exact data experiment: iterates for $N_t = N_r = 100$ with $N = 1$ (left), $N = 5$ (middle) and $N = 10$ (right) after 1, 5 and 25 cycles.

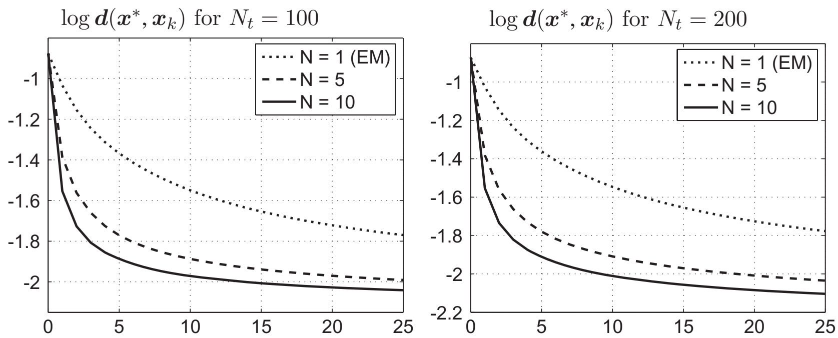

Figure 3. Exact data experiment: logarithmic plots of iteration error with respect to the Kullback–Leibler distance for $N_t = N_r = 100$ (left) and $N_t = N_r = 200$ (right).

dependence of the noise level, with $\tau(\delta) < 1$ for large δ and $\tau(\delta)$ converging to some $\tau_\infty > 1$ for $\delta \to 0$. It is clear that the asymptotic convergence analysis (for $\delta \to 0$) remains valid in such a situation.

The reconstruction for noisy data with $N_t = N_r = 100$ is depicted in figure 4. For comparison purposes, results of the OS-EM iteration (without loping strategy) are also included. The loping OS-EM is automatically stopped according to (4.9) whereas their non-loping counterparts are stopped after the cycle with minimal error $\boldsymbol{d}\big(\boldsymbol{x}^*, \boldsymbol{x}_k^\delta\big)$, which is not

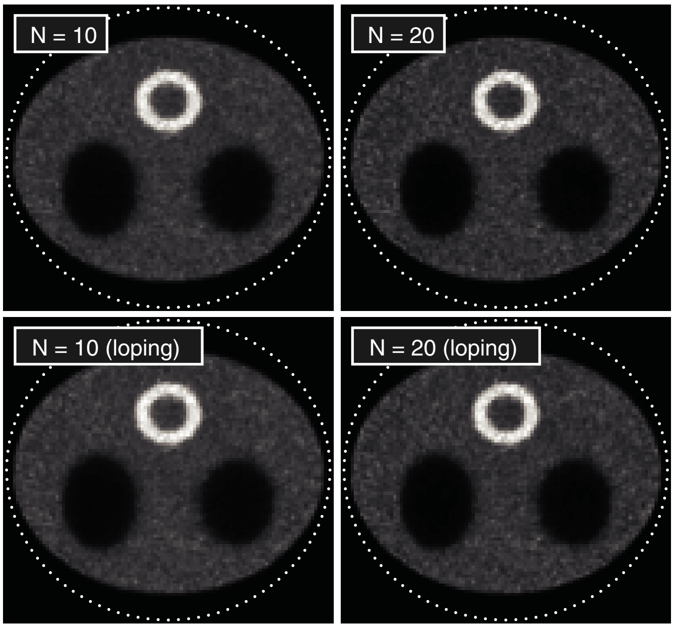

Figure 4. Noisy data experiment: iterates without loping (top line) and with loping (bottom line). The loping iterations are stopped automatically whereas their non-loping counterparts are stopped at the iteration cycle where $d(x^*, x_k^\delta)$ is minimal (which is not available in practice).

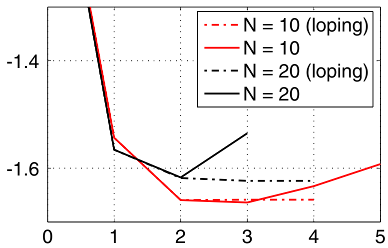

Figure 5. Noisy data experiment: evolution of the relative error $\log \boldsymbol{d}(\boldsymbol{x}^*, \boldsymbol{x}_k^\delta)$ for loping and non-loping OS-EM iterations.

Table 1. Comparison of the performance of different iterative methods. The non-loping iterations are stopped after the cycle with minimal error, whereas the loping OS-EM are automatically stopped according to (4.9).

	N	N_{cycl}	Time (s)	$\boldsymbol{d}(\boldsymbol{x}^*, \boldsymbol{x}_k^\delta)$
Loping OS-EM	10	4	13.4	0.022
OS-EM	10	3	9.2	0.022
Loping OS-EM	20	4	13.4	0.024
OS-EM	20	2	6.3	0.024

available in practice. All reconstructions are quite comparable. Figure 5 shows the evolution of the error with respect to the KL-distance. In this figure one also notes the semi-convergence of the non-loping iterations, which happens typically when applying non-regularized iterative schemes to ill-posed problems [4, 11, 14].

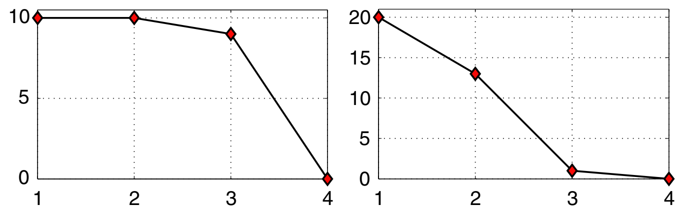

Figure 6. Noisy data experiment: the x-axis shows the number of cycles, while the number of actually performed iterations within each cycle is shown at the y-axis.

An inspection of figure 5 shows that the regularized solution of the loping OS-EM methods (automatically stopped) have errors comparable to the optimal solution of their non-loping counterparts when stopped after the cycle with minimal error (which is not available in the practice). Figure 6 shows the number of actually performed iterations. Table 1 summarizes run times and errors with $N_t = N_r = 100$, $N_{\text{angle}} = 100$ (with non-optimized Matlab implementation on HP Notebook with 2 GHz Intel Core Duo processor).

6. Conclusions

This paper is devoted to the investigation of OS-EM type algorithms for solving systems of linear ill-posed equations. We focus on showing regularization properties of the proposed methods.

In the case of exact data, our approach yields an algorithm analog to the OS-EM iteration. We are able to prove monotonicity results with respect to the Kullback–Leibler distance as well as weak convergence in the case of boundedness of the iterations. In the noisy data case, we propose a loping OS-EM iteration which differs from the OS-EM method due to the introduction of a loping strategy. This loping strategy renders the proposed iteration of a regularization method. We prove monotonicity of the iterates and study stability properties of our method.

The boundedness away from zero and the normalization conditions on which this work is based are first steps toward providing stopping rules for EM-type algorithms used in PET. Relaxing the assumptions is a further research subject.

What concerns the numerical effort, we conjecture that the loping OS-EM algorithm is at least as efficient as the well-established OS-EM method. The numerical experiments with (5.5) for inverting the circular Radon transform support this conjecture. In the case of exact data, (5.5) reduces to a discretized version of the continuous OS-EM iteration applied to system (5.3). However it is slightly different to the discrete OS-EM iteration of [12] since $\mathbf{B}_j$ is not the exact transpose of $\mathbf{M}_j$. Moreover, opposed to [12] our continuous convergence analysis applies independent on the discretization level.

Acknowledgments

MH has been supported by the technology transfer office of the University Innsbruck (transIT) within the framework of the NFN 'Photoacoustic Imaging in Biology and Medicine' from the Austrian Science Foundation (project S10505-N20). The work of AL is supported by the Brazilian National Research Council CNPq, grants 306020/2006-8, 474593/2007-0, and by the Alexander von Humbolt Foundation AvH. ER acknowledges support from the Austrian Science Foundation, Elise Richter scholarship (V82-N18 FWF).

References

[1] Byrne C 2009 Block-iterative algorithms *Int. Trans. Oper. Res.* to appear

[2] De Cesaro 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

[3] Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc.* B **39** 1–38

[4] Deuflhard P, Engl H W and Scherzer O 1998 A convergence analysis of iterative methods for the solution of nonlinear ill–posed problems under affinely invariant conditions *Inverse Problems* **14** 1081–106

[5] Eggermont P P B and LaRiccia V N 1996 Maximum penalized likelihood estimation and smoothed EM algorithms for positive integral equations of the first kind *Numer. Funct. Anal. Optim.* **17** 737–54

[6] Eggermont P P B 1999 Nonlinear smoothing and the EM algorithm for positive integral equations of the first kind *Appl. Math. Optim.* **39** 75–91

[7] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems (Mathematics and its Applications* vol 375*)* (Dordrecht: Kluwer)

[8] Fessler J A, Clinthorne N H and Rogers W L 1993 On complete data spaces for PET reconstruction algorithms *IEEE Trans. Nucl. Sci.* **40** 1055–61

[9] 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

[10] 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

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

[12] Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets projection data *IEEE Trans. Med. Imaging* **13** 601–9

[13] Iusem A N 1992 A short convergence proof of the EM algorithm for a specific Poisson model *Rebrape* **6** 57–67

[14] 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)

[15] Kondor A 1983 Method of convergent weights: an iterative procedure for solving Fredholm's integral equations of the first kind *Nucl. Instrum. Methods Phys. Res.* **216** 177–81

[16] Kowar R and Scherzer O 2002 Convergence analysis of a Landweber–Kaczmarz method for solving nonlinear ill-posed problems *Ill-Posed Inverse Problems* **23** 69–90

[17] Kuchment P and Kunyansky L A 2008 Mathematics of thermoacoustic and photoacoustic tomography *Eur. J. Appl. Math.* **19** 191–224

[18] Lucy L B 1974 An iterative technique for the rectification of observed distributions *Astron. J.* **79** 745–54

[19] Mair B A, Rao M and Anderson J M M 1996 Positron emission tomography, Borel measures and weak convergence *Inverse Problems* **12** 965–76

[20] Mülthei H N and Schorr B 1987 On an iterative method for a class of integral equations of the first kind *Math. Methods Appl. Sci.* **9** 137–68

[21] Mülthei H N and Schorr B 1989 On properties of the iterative maximum likelihood reconstruction method *Math. Methods Appl. Sci.* **11** 331–42

[22] Mülthei H N 1993 Iterative continuous maximum-likelihood reconstruction methods *Math. Methods Appl. Sci.* **15** 275–86

[23] Paltauf G, Nuster R, Haltmeier M and Burgholzer P 2007 Experimental evaluation of reconstruction algorithms for limited view photoacoustic tomography with line detectors *Inverse Problems* **23** 81–94

[24] Redner R A and Walker H F 1984 Mixture densities, maximum likelihood and the EM algorithm *SIAM Rev.* **26** 195–239

[25] Resmerita E and Anderssen R S 2007 Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems *Math. Methods Appl. Sci.* **30** 1527–44

[26] Resmerita E, Engl H W and Iusem A N 2007 The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis *Inverse Problems* **23** 2575–88

[27] Richardson W H 1972 Bayesian-based iterative method of image restoration *J. Opt. Soc. Am.* **62** 55–9

[28] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging (Applied Mathematical Sciences* vol 167*)* (New York: Springer)

[29] Vardi Y, Shepp L A and Kaufman L 1985 A statistical model for positron emission tomography *J. Am. Stat. Assoc.* **80** 8–37 (with discussion)

[30] Xu M and Wang L V 2006 Photoacoustic imaging in biomedicine *Rev. Sci. Instrum.* **77** 041101