Intl. Trans. in Op. Res. 15 (2008) 417–438

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Incomplete oblique projections method for solving regularized least-squares problems in image reconstruction

H. D. Scolnik[a], N. E. Echebest[b] and M. T. Guardarucci[c]

[a]*Departamento de Computación, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina,*
[b]*Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina,*
[c]*Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata, Argentina*
E-mail: hugo@dc.uba.ar [Scolnik];opti@mate.unlp.edu.ar [Echebest]; marite@mate.unlp.edu.ar [Guardarucci]

Received 14 September 2007; received in revised form 13 February 2008; accepted 16 February 2008

Abstract

In this paper we improve on the incomplete oblique projections (IOP) method introduced previously by the authors for solving inconsistent linear systems, when applied to image reconstruction problems. That method uses IOP onto the set of solutions of the augmented system $Ax - r = b$, and converges to a weighted least-squares solution of the system $Ax = b$. In image reconstruction problems, systems are usually inconsistent and very often rank-deficient because of the underlying discretized model. Here we have considered a regularized least-squares objective function that can be used in many ways such as incorporating blobs or nearest-neighbor interactions among adjacent pixels, aiming at smoothing the image. Thus, the oblique incomplete projections algorithm has been modified for solving this regularized model. The theoretical properties of the new algorithm are analyzed and numerical experiments are presented showing that the new approach improves the quality of the reconstructed images.

Keywords: least-squares problems; minimum norm solution; regularization; image reconstruction; computerized tomography; incomplete projections

1. Introduction

Large and sparse systems of linear equations arise in many important applications (Censor and Zenios, 1997), such as radiation therapy treatment planning, computational mechanics, and optimization, and in image processing problems such as electromagnetic geotomography (Popa and Zdunek, 2004). In practice, most problems arising from tomographic image reconstructions are inconsistent and of deficient rank.

Many problems in the field of tomographic image reconstruction are modeled by the linear least-squares problem, that is: find $x^* \in \Re^n$ such that

$$\min_x \|Ax - b\|^2_{D_m}, \tag{1}$$

Published by Blackwell Publishing, 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA.

where A is an $m \times n$ matrix, $b \in \Re^m$, $\|.\|_{D_m}$ denotes a weighted norm, and D_m is a positive-definite matrix.

A widely used algorithm in Computerized Tomography is ART (Algebraic Reconstruction Technique), whose origin goes back to Kaczmarz (Kaczmarz, 1937; Censor and Zenios, 1997), although it is known that in order to obtain convergence in the inconsistent case it is necessary to use an underrelaxation parameter that must tend to zero. Popa has developed an extension of ART called KERP (Popa and Zdunek, 2004) that converges for inconsistent systems, and more recently in Popa and Zdunek (2005) the authors showed its efficiency in the case of rank-deficient systems. For the classic simultaneous projection algorithms (Censor and Zenios, 1997) for inconsistent problems (Landweber, 1951), CAV (Censor et al., 2001), and the more general class of the diagonal weighting (DWE) algorithms (Censor and Elfving, 2002), in order to prove convergence it is necessary to choose the relaxation parameter in an interval that depends on the largest eigenvalue of A^TA (Byrne, 2002; Jiang and Wang, 2003). Within the framework of the Projected Aggregation Methods (PAM) (García Palomares, 1993), we have developed acceleration schemes based on projecting the search directions onto the aggregated hyperplanes, with excellent results in both consistent and inconsistent systems (Scolnik et al., 2002a, 2008; Echebest et al., 2005).

In Scolnik et al. (2008), we introduced the IOP algorithm for solving linear least-squares problems that uses a scheme of incomplete oblique projections (IOP) onto the solution set of the augmented system $Ax - r = b$, and converges to a weighted least-squares solution of the system $Ax = b$ when $rank\ (A) = n$. Also, we extended those results in order to compute the weighted least-squares solution of inconsistent and rank-deficient systems (Scolnik et al., 2006), proving convergence to the minimum norm solution.

The tomographic image reconstruction problems are such that the limitation of the range of rays makes the model underdetermined, the discretized linear system is rank-deficient and ill-conditioned, the nullspace is non-trivial, and the minimal norm least-squares solution may be far away from the true image. When methods that converge to the minimum norm solution are used, in order to recover missing components several authors considered a regularized weighted least-squares problem (Björk, 1996; Tikhonov, 1963; Popa and Zdunek, 2005):

$$\min_{x \in \Re^n} \|Ax - b\|^2_{D_m} + \beta R(x), \tag{2}$$

where D_m is a matrix of weights of data, β is a regularization parameter, and $R(x)$ is a function that enforces smoothness in the image.

The term $R(x)$ can be defined in many forms. In particular, the Tikhonov regularization (Tikhonov, 1963; Hansen, 1998) in its standard form is defined by a square of discrete smoothing norm $\|Mx\|^2$, where M (called the regularization operator) is typically either an identity or a diagonal matrix that defines a norm or semi-norm on the solution. The regularization parameter β is a positive constant chosen to control the size of the solution vector. As explained for instance by Popa and Zdunek (2005), and Pralat and Zdunek (2005), $R(x)$ can be interpreted in terms of the Gibbs prior (Lange, 1990) that is used in statistical image reconstruction. Koltracht et al. (1990) proved that the image components in $N(A)$ have a pattern of vertical stripes, and this is visible in the image of minimal-norm solution as vertical smearing from inhomogeneous features. This effect can be partially reduced if interactions among neighboring pixels are incorporated into the reconstruction problem. This idea leads to define the discrete smoothing norm in the

regularized function (see Popa and Zdunek, 2005; Lange, 1990) as $R(x) = 2U(x)$, where

$$U(x) = \sum_{j=1}^{n} \sum_{i \in S_j} w_{ji} V(x_j - x_i, \delta),$$

where S_j is a set of indices of the nearest neighborhood of pixel j, w_{ji} a factor of weight, and $V(x_j - x_i, \delta)$ a potential function, with δ being a scaling parameter. There are several proposals aimed at the same objective (Lange, 1990), and among them we chose $V(x_j - x_i, \delta) = \left(\frac{x_j - x_i}{\delta}\right)^2$. This gives rise to the problem

$$\min_{x \in \Re^n} \frac{1}{2} \|Ax - b\|_{D_m}^2 + \frac{2\beta}{\delta^2} x^T M x, \tag{3}$$

where the matrix M is symmetric and semi-positive definite, depending only on the weights w_{ji} and the S_j sets of indices of the nearest neighborhood of each pixel j. This model can be described by

$$\min_{x \in \Re^n} \frac{1}{2} \|Ax - b\|_{D_m}^2 + \frac{1}{2} \|M^{1/2} x\|_{D_n}^2, \tag{4}$$

considering the spectral decomposition of $M = U^T \Lambda U = M^{1/2^T} M^{1/2}$, and the diagonal D_n with a number $\alpha = \frac{4\beta}{\delta^2}$ of non-zero elements.

In this paper we have replaced in representation (3) the minimization of the convex function $x^T M x$ by the minimization of the D_n norm of its gradient, $\|Mx\|_{D_n}^2$, because they are theoretically equivalent. Thus, the regularized least-squares model we consider is

$$\min_{x \in \Re^n} \frac{1}{2} \|Ax - b\|_{D_m}^2 + \frac{1}{2} \|Mx\|_{D_n}^2, \tag{5}$$

where D_n is the matrix defined previously.

The regularization parameter β in (2) or the diagonal of D_n in (5) controls the degree of regularization applied to the original problem, although it is not usually known beforehand. Depending on the problem and the solver being used, several methods can be used to choose an appropiate regularization parameter. Two such methods are the generalized cross-validation (GCV) criterion (Wahba, 1977, 1990; Björk, 1996) and the L-curve criterion (Hansen and O'Leary, 1993; Hansen, 1998).

In this paper, we present a version of the IOP algorithm (Scolnik et al., 2006) for solving regularized least-squares problems (5). This algorithm, called regularized IOP (RIOP), adopts the basic IOP scheme described below as Algorithm 1, but allowing many possibilities for defining both the regularization matrix M and the norms D_n, to deal with different sorts of problems. In order to solve (5), we define two convex sets in the $(2n+m)$-dimensional space $\Re^{2n+m}$, denoting by $[u; v]$ the vertical concatenation of $u \in \Re^n$, with $v \in \Re^{m+n}$,

$$\mathscr{P} = \{p : p = [x; r], x \in \Re^n, r = [r_1; r_2] \in \Re^{m+n}, \text{such that } Ax - r_1 = b, Mx - r_2 = 0\}, \text{and} \tag{6}$$

$$\mathscr{Q} = \{q : q = [x; 0], x \in \Re^n, 0 \in \Re^{m+n}\}, \tag{7}$$

$d(p,q) = \|p-q\|_D$, for all $p \in \mathscr{P}$, $q \in \mathscr{Q}$. D is a diagonal matrix of order $2n+m$, whose n first elements are 1's, the next m coincide with those of D_m, and the last n elements are those of D_n.

By means of a direct application of the Karush–Kuhn–Tucker (KKT) conditions (Luenberger, 1986) to the problem

$$\min\{\|p-q\|_D^2 : \text{for all } p \in \mathscr{P} \quad \text{and} \quad q \in \mathscr{Q}\}, \tag{8}$$

it is possible to prove (see Scolnik et al., 2008) that this is equivalent to (5). This observation led us to modify the IOP algorithm for solving (5), applying an alternate projections scheme between the sets $\mathscr{P}$ and $\mathscr{Q}$, as in the original development in Scolnik et al. (2008). This procedure is similar to the one of Csiszár and Tusnády (1984), but replacing the computation of the exact projections onto $\mathscr{P}$ by suitable incomplete or approximate projections, according to the following basic scheme:

Algorithm 1 (*Basic Alternating Scheme*)
Iterative step: *Given* $p^k = [x^k; r^k] \in \mathscr{P}$, $q^k = [x^k; 0] \in \mathscr{Q}$, find $p_a^{k+1} = [x^{k+1}; r^{k+1}] \in \mathscr{P}$ as $p_a^{k+1} \approx \arg\min\{\|p-q^k\|_D^2 : p \in \mathscr{P}\}$, then define $p^{k+1} = p_a^{k+1}$, and $q^{k+1} \in \mathscr{Q}$ means of $q^{k+1} = [x^{k+1}; 0] \equiv \arg\min\{\|p^{k+1}-q\|_D^2 : \quad q \in \mathscr{Q}\}$.

In order to compute the incomplete projections onto $\mathscr{P}$ we apply our ACCIM algorithm (Scolnik et al., 2002b, 2008). It is convenient to point out, for clarifying the applicability of ACCIM within the new approach, that given a consistent system $\bar{A}y = \bar{b}$, the sequence $\{y^k\}$ generated by ACCIM from the initial point y^0 converges to the solution $\bar{y}^*$ of $\bar{A}y = \bar{b}$, satisfying $\bar{y}^* = \arg\min\{\|y^* - y^0\|_D^2, \ y^* \in \Re^n : \bar{A}y^* = \bar{b}\}$.

The sequence $\{y^k\}$ is Fejér monotone with respect to the solution set of $\bar{A}y = \bar{b}$ (Scolnik et al., 2002b). This iterative algorithm uses simultaneous projections onto the hyperplanes defined by the system of equations of $\mathscr{P}$, and is very efficient for solving consistent problems and is convenient for computing approximate projections with some required properties, as explained in Scolnik et al. (2008). In this paper, and in a more general context than the one in Tikhonov (1963), the L-curve is used to determine the feasible values of the regularization parameter for all problems.

In the following sections we will present the RIOP algorithm, the criterion for accepting approximate projections, together with some related results needed for the corresponding convergence theory.

The test problems are two simulated reconstruction problems in borehole electromagnetic geotomography appearing in Popa and Zdunek (2005), and two in computerized tomography from SNARK (Browne et al., 1993).

2. Projection algorithms

Hereafter $\|x\|$ will denote the *Euclidean* norm of $x \in \Re^n$, and $\|x\|_D$ the norm induced by a positive-definite matrix D. We will also assume that each row of A has a *Euclidean* norm equal to 1. We will use the notation $\mathbf{e}_i$ for the i-th column of I_n, where the symbol I_n denotes the identity matrix in $\Re^{n\times n}$, and the upper index T for the transpose of a matrix. Given $W \in \Re^{n\times r}$, we will denote by $\mathbf{w}_i^T$ the i-th row of W and by $R(W)$ the subspace spanned by the columns of W, P_W and P_W^D being the orthogonal and the oblique projectors onto $R(W)$. We will use the notation $R(W)^{\perp}$ for the

D-orthogonal subspace to $R(W)$, and by $P^D_{W^\perp}$ the corresponding projector. In particular, if $W = [v] \in \Re^{n\times 1}$, we will use $P^D_{v^\perp}$. We denote a diagonal matrix of order n by $D = diag(d)$, where $d = (d_1, \ldots, d_n)$.

Let us assume we have a compatible system $\bar{A}y = \bar{b}$, $\bar{A} \in \Re^{m\times n}$, $m \geqslant n$, $\bar{b} \in \Re^m$. For each constraint of $\bar{A}y = \bar{b}$, we will denote by $L_i = \{y \in \Re^n : \bar{\mathbf{a}}_i^T y = b_i\}$, $r_i(y) = \bar{\mathbf{a}}_i^T y - b_i$, and the oblique projection of y onto L_i by

$$P_i^D(y) = y - \frac{r_i(y)}{\bar{\mathbf{a}}_i^T D^{-1} \bar{\mathbf{a}}_i} D^{-1} \bar{\mathbf{a}}_i . \tag{9}$$

3. IOP algorithm for the regularized least-squares problem

In order to solve the regularized weighted least-squares problem (5), we consider its equivalence with (8). Hence, we will apply the alternate projections scheme (1) between the sets $\mathscr{P}$ and $\mathscr{Q}$, defining the conditions for accepting an incomplete or an approximate projection onto $\mathscr{P}$.

We consider a diagonal weighting matrix $D \in \Re^{N\times N}$, where $N = 2n+m$, such that

$$D = \begin{pmatrix} I_n & 0 & 0 \\ 0 & D_m & 0 \\ 0 & 0 & D_n \end{pmatrix}, \tag{10}$$

and the sets $\mathscr{P}$ and $\mathscr{Q}$ as described in (6) and (7). The diagonal matrices D_m and D_n are defined in such a way that it allows to apply weights to the residuals r_1 and r_2.

Given $p^k \in \mathscr{P}$, and its projection q^k onto $\mathscr{Q}$, we will denote by $p^D_{min}(q^k)$ the projection of q^k onto $\mathscr{P}$, which is the solution to the problem

$$\min\{\|p - q^k\|_D : \ p \in \mathscr{P}\}. \tag{11}$$

Given $q^k \in \mathscr{Q}$, instead of defining $p^{k+1} = p^D_{min}(q^k)$, we define $p^{k+1} = p_a^{k+1}$, where $p_a^{k+1} \in \mathscr{P}$ is a point obtained by means of the incomplete resolution of problem (11).

Remark 1 *Algorithm 1 based on exact projections is always convergent (Byrne and Censor, 2001), but its computational cost is high.*

We have presented in Scolnik et al. (2008) the theory of inexact projections aimed at obtaining similar convergence properties but with a much lower computational cost.

In order to define the inexact projection $p_a^{k+1} \approx p^D_{\min}(q^k)$, we consider the following:

Definition 1 *Given an approximation* $\hat{p} = [z; \mu]$ of $p^D_{\min}(q^k)$, *where* $z \in \Re^n$, $\mu = [\mu_1; \mu_2] \in \Re^{m+n}$ *satisfy* $s_1 = Az - \mu_1 - b$ and $s_2 = Mz - \mu_2$, *we will denote by* $P(\hat{p}) = [z; \mu + s]$ *the solution of the system* $Az - r_1 = b$, $Mz - r_2 = 0$, *such that* $[r_1; r_2] = \mu + s = [\mu_1 + s_1; \mu_2 + s_2]$.

Aiming at obtaining properties of the sequence $\{p^k\}$ generated by the new algorithm that guarantees convergence to the solution of (8), we establish an "acceptance condition" that an approximation $\hat{p} = [z; \mu]$ of $p^D_{\min}(q^k)$ must satisfy.

When applying ACCIM for obtaining the approximate projection p_a^{k+1} we generate, according to Scolnik et al. (2002a, 2008), a sequence $\{\hat{p}^j\}$, with initial point $[q^k, 0]$, that is Fejér monotone

with respect to $p^D_{\min}(q^k)$. That is, it satisfies $\|\hat{p}^{j+1} - p^D_{\min}(q^k)\|^2_D \leqslant \|\hat{p}^j - p^D_{\min}(q^k)\|^2_D$. Furthermore, if $p^D_{\min}(q^k) \neq p^k$, $\|\hat{p}^j - p^D_{\min}(q^k)\|^2_D < \|\hat{p}^j - p^k\|^2_D$.

Now, considering that the sequence generated by ACCIM is Fejér monotone, we can define in an easier way than in Scolnik et al. (2006, 2008) the following condition.

Definition 2. *Acceptance Condition. An approximation* $\hat{p}^j = [z^j; \mu^j]$ of $p^D_{\min}(q^k)$, *using ACCIM, is acceptable if*

$$\|\hat{p}^j - P(\hat{p}^j)\|^2_D \leqslant \gamma \|\hat{p}^j - p^k\|^2_D, \quad \text{with } 0 < \gamma < 1 \tag{12}$$

holds.

Remark 2. *In particular,* $\hat{p} = p^D_{\min}(q^k)$ *satisfies* (12).

We have proved in Lemma 1 in Scolnik et al. (2008) that using the ACCIM Algorithm, and taking into account the properties of the generated sequence, it is possible to find a $j^* > 0$, *such that* $[z^{j^*}; \mu^{j^*}]$ *satisfies* (12). *In particular, if* $p^D_{\min}(q^k) = p^k$, *then* $p^k = \hat{p}^1$.

In order to describe the alternate incomplete projections algorithm, we define the approximation p^{k+1}_a of $p^D_{\min}(q^k)$, by the following.

Definition 3.

$$p^{k+1}_a = P(\hat{p}^{j^*}), \quad \text{if } \hat{p}^{j^*} = [z^{j^*}; \mu^{j^*}] \text{ satisfies (12).} \tag{13}$$

We present in the following a practical algorithm to solve (8).

Algorithm 2. *Regularized Incomplete Oblique Projections (RIOP)*

Initialization: Given $0 < \gamma \leqslant 1/2$, a positive-definite diagonal matrix D_m of order m, a matrix D_n of order n, and $p^0 = [x^0; r^0]$, set $r^0 = [Ax^0 - b; Mx^0]$, $q^0 = [x^0; 0] \in \mathcal{Q}$ and $k \leftarrow 0$.

Iterative Step: *Given* $p^k = [x^k; r^k]$ *and* $q^k = [x^k; 0]$.

- *Calculate* $\hat{p}^*$, *an approximation of* $p^D_{\min}(q^k)$ *satisfying* (12), *applying ACCIM as follows: define* $y^0 = [z^0; \mu^0] = q^k$ *the initial point.*
- *For solving* $Ax - r_1 = b$, $Mx - r_2 = 0$ *iterate until finding* $y^j = [z^j; \mu^j]$, *with* $\mu^j = [{\mu_1}^j; {\mu_2}^j] \in \Re^{m+n}$ *such that* ${s_1}^j = Az^j - {\mu_1}^j - b$ *and* ${s_2}^j = Mz^j - {\mu_2}^j$ *satisfy* (12), *that is*

$$\|{s_1}^j\|^2_{D_m} + \|{s_2}^j\|^2_{D_n} \leqslant \gamma\left(\|r^k_1\|^2_{D_m} + \|r^k_2\|^2_{D_n} - S_j\right), \quad \text{with} \quad S_j = \|y^j - y^0\|^2_D. \tag{14}$$

- *Define* $p^{k+1} = [x^{k+1}; r^{k+1}]$, *with* $x^{k+1} = z^j$, *and* $r^{k+1} = \mu^j + s^j$.
- *Define* $q^{k+1} = [x^{k+1}; 0] \in \mathcal{Q}$.
- $k \leftarrow k+1$.

Aiming at solving rank-deficient problems as in Scolnik et al. (2006), we assume in algorithm 2 that the parameter γ is restricted to $0 < \gamma \leqslant 1/2$. This hypothesis allowed us to extend in Scolnik et al. (2006) the convergence results given in Lemma 4 of Scolnik et al. (2008) to problems with rank $(A) < n$. The same convergence results can be applied to the RIOP algorithm as explained next.

4. Convergence of the RIOP algorithm

Problem (5), considering that the matrix is vertically concatenated $\tilde{A} = [A; M]$, $\tilde{b} = [b; 0] \in \Re^{m+n}$ and that the diagonal matrix $D_{n+m} \in \Re^{m \times n}$ is formed with the elements of D_m and D_n, can be stated as

$$\min_{x \in \Re^n} \|\tilde{A}x - \tilde{b}\|^2_{D_{m+n}}. \tag{15}$$

We consider the set:

$$L^{sq}_{D_{m+n}} = \{x^* \in \Re^n : \text{ for which } r^* = \tilde{A}x^* - \tilde{b} \text{ satisfies } \tilde{A}^T D_{m+n} r^* = 0\}, \tag{16}$$

that is the set of solutions to the problem (8), and the corresponding

$$S_{D_{m+n}} = \{p^* : p^* = [x^*; r^*] \in P \text{ such that } x^* \in L^{sq}_{D_{m+n}}\}. \tag{17}$$

Furthermore, given as the initial point $q^0 = [x^0; 0] \in \mathcal{Q}$, we denote

$$\bar{x}^* = \arg \min_{\{x^* \in L^{sq}_{D_{m+n}}\}} \left\|x^0 - x^*\right\|^2 \tag{18}$$

and $\bar{p}^* = [\bar{x}^*; r^*] \in S_{D_{m+n}}$. This point satifies

$$\bar{p}^* = \arg \min_{\{p^* \in S_{D_{m+n}}\}} \left\|q^0 - p^*\right\|^2_D, \tag{19}$$

because $\|q^0 - p^*\|^2_D = \|x^0 - x^*\|^2 + \|r^*\|^2_{D_{m+n}}$.

The results in Scolnik et al. (2008), Lemma 2 and Lemma 3, allow us to prove that the sequence $\{p^k\}$ generated by the new version of the IOP algorithm is convergent to p^* in the following theorem.

Theorem 1. *Let $\{p^k\}$ be the sequence generated by Algorithm 2, using $0 < \gamma \leqslant 1/2$. If $\bar{p}^* = [\bar{x}^*; r^*]$ is the element defined in* (19)*, and its projection $\bar{q}^* = [\bar{x}^*; 0] \in \mathcal{Q}$, then*

(i) *The sequence $\left\{\left\|p^k - \bar{q}^*\right\|^2_D\right\}$ is decreasing and bounded, and hence it converges.*
(ii) *The sequence $\{p^k\}$ converges to $\bar{p}^*$.*

With the purpose of analyzing the algorithm's behavior, we need to describe some properties related to the inner iterative steps arising from the use of ACCIM, which is the basis for computing approximate solutions to problem (11) in the RIOP algorithm. These results are needed for proving the convergence of the new algorithm.

4.1. Properties of the ACCIM algorithm

Assume that $\bar{A}y = \bar{b}$ is a consistent system where $\bar{A} \in \Re^{m \times n}$, and y^* a solution to it. Let $\{y^j\}$ be the sequence generated by a version of the ACCIM algorithm with a D-norm, and $s^j = \bar{A}y^j - \bar{b}$ the residual at each iterate y^j.

The direction d^j defined in ACCIM (see Appendix A in Scolnik et al., 2008) by combining projections (9) is

$$d^j = \sum_{l=1}^{m} w_l (P_l^D(y^j) - y^j) = -\sum_{l=1}^{m} w_l \frac{s_l{}^j}{\|\bar{a}_l\|^2_{D^{-1}}} D^{-1} a_l. \tag{20}$$

At each iterate $y^j \neq y^*$, $j>0$, the direction used is $\hat{d}^j = P_{v^\perp}(d^j)$, where $v = \hat{d}^{j-1}$, and the next iterate $y^{j+1} = y^j + \lambda_j \hat{d}^j$ is defined for satisfying

$$(\hat{d}^j)^T D(y^j + \lambda_j \hat{d}^j - y^*) = 0. \tag{21}$$

Furthermore, from the definition of $\hat{d}^j$ and λ_j, it is possible to obtain:

Lemma 1. *If $y^j \neq y^*$, $j>0$, is generated by the ACCIM algorithm, then*

(i) d^j is D-orthogonal to $y^* - y^i$, for all $i<j$.
(ii) $\hat{d}^j$ is D-orthogonal to $\hat{d}^i$ and d^i, for all $i<j$.

Furthermore,

(iii) $y^* - y^j$ is D-orthogonal to $\hat{d}^i$, for all $i<j$, and as a consequence is also D-orthogonal to $y^j - y^0$.

Proof. See the proof of Lemma 2 in Appendix A of Scolnik et al. (2008).

In particular, the application of ACCIM for solving $\tilde{A}z - I_{m+n}r = \tilde{b}$ has the following characteristics:

Given $p^k = [x^k; r^k]$ and $q^k = [x^k; 0]$, $k \geqslant 0$, ACCIM computes an approximation $[z^j; \mu^j]$ to the projection $p^D_{\min}(q^k)$.

From the starting point $[z^0; \mu^0] = q^k = [x^k; 0]$, for each iterate $[z^j; \mu^j]$, we denote by s^j the residual $s^j = \tilde{A}z^j - \mu^j - \tilde{b}$. The direction $d^j \in \Re^{n+m+n}$ can be written as

$$d^j = -\sum_{i=1}^{m+n} w_i \frac{s_i^j}{\|\bar{\mathbf{a}}_i\|^2_{D^{-1}}} D^{-1} \bar{\mathbf{a}}_i, \tag{22}$$

$\bar{\mathbf{a}}_i = [\mathbf{a}_i; -\mathbf{e}_i]$ being the i-th column of $[\tilde{A}, -I_{m+n}]^T$, where $\mathbf{e}_i$ denotes the i-th column of I_{m+n}. The square of the norm of each row of the matrix $[\tilde{A}, -I_{m+n}]$, induced by the inverse of D, is $1+[1/(D_{m+n})_i]$, if $\|\mathbf{a}_i\| = 1$, and $(D_{m+n})_i$ denotes the i-th element of the diagonal of D_{m+n}. Hence, the direction $d^j = [d^j_1; d^j_2]$ has its first n components given by

$$d^j_1 = -\sum_{i=1}^{m+n} \frac{w_i (D_{m+n})_i s_i^j}{1 + (D_{m+n})_i} \mathbf{a}_i \tag{23}$$

and the next $m+n$ components are

$$d^j_2 = (D_{m+n})^{-1} \sum_{i=1}^{m+n} \frac{w_i (D_{m+n})_i s_i^j}{1 + (D_{m+n})_i} \mathbf{e}_i. \tag{24}$$

We choose for this application of ACCIM the values $w_i = 1 + (D_{m+n})_i$, because they privilege the rows with greatest $1+(D_{m+n})_i$. Thus, at each iterate $[z^j; \mu^j]$, the directions are

$$d_1{}^j = -\sum_{i=1}^{m+n} (D_{m+n})_i \mathbf{a}_i s_i{}^j = \tilde{A}^T D_{m+n}(-s^j), \text{ and } d_2{}^j = \sum_{i=1}^{m+n} \mathbf{e}_i s_i{}^j = s^j. \tag{25}$$

This direction $\hat{d}^j$ is such that $d_1{}^j = \tilde{A}^T D_{m+n}(-d_2{}^j)$. Similar behavior has the direction $d^j = [\hat{d}_1{}^j; \hat{d}_2{}^j]$, where $\hat{d}_1{}^j = \tilde{A}^T D_{m+n}(-\hat{d}_2{}^j)$. This expression follows from considering $\hat{d}_1{}^j = d_1{}^j - \dfrac{(\hat{d}^{j-1})^T D d^j}{\|\hat{d}^{j-1}\|_D^2}\hat{d}_1{}^{j-1}$, $\hat{d}_2{}^j = d_2{}^j - \dfrac{(\hat{d}^{j-1})^T D d^j}{\|\hat{d}^{j-1}\|_D^2}\hat{d}_2{}^{j-1}$, and using that $\hat{d}^0$ coincides with d^0. Hence, for each $j>0$, $[z^j; \mu^j]$ satisfies $z^j - z^0 = \tilde{A}^T D_{m+n}(-\mu^j)$.

Remark 3. *We will now describe the properties of the accepted approximation $\hat{p} = [z^j; \mu^j]$, in the iterations of Algorithm 2.*

(a) From (iii) of the previous Lemma, we know that for every solution $[z^*; \mu^*]$ of $\tilde{A}z - \mu = \tilde{b}$, $[z^*; \mu^*] - [z_j; \mu_j]$ is D-orthogonal to $[z_j; \mu_j] - [z_0, \mu_0]$. Thus, for all $p \in \mathscr{P}$, we have that

$$(p - \hat{p})^T D(\hat{p} - q^k) = 0. \tag{26}$$

(b) From (i) of the previous Lemma, we know that for every $[z^j; \mu^j], j > 0$, the direction d^j satisfies $d^{jT} D([z^*; \mu^*] - [z^0; \mu^0]) = 0$. Then, because $p^k \in \mathscr{P}$ and $q^k = [z^0; \mu^0]$, in particular $d^{jT} D(p^k - q^k) = 0$. Hence, using the expression of d^j in (25), and considering that $p^k - q^k = [0; r^k]$, we obtain

$$s^j \text{ is } D_{m+n}\text{-orthogonal to } r^k. \tag{27}$$

(c) Furthermore, considering (a) and (b), it follows that $p^k - \hat{p}$ is D-orthogonal to $[0; s^j]$.

Remark 4. *The inequality* (14)*, used in Algorithm 2 to accept $[z^j; \mu^j]$, arises from replacing in the inequality of* (12)*, $\|\hat{p} - P(\hat{p})\|_D^2 = \|s^j\|_{D_{m+n}}^2$, and $\|p^k - \hat{p}\|_D^2 = \|(p^k - q^k) - (\hat{p} - q^k)\|^2$. Moreover, from* (26) *we know that $p^k - \hat{p}$ and $\hat{p} - q^k$ are D-orthogonal, and using the fact that $\|p^k - q^k\|_D^2 = \|[0; r^k]\|_D^2 = \|r^k\|_{D_{m+n}}^2$, we obtain $\|p^k - \hat{p}\|_D^2 = \|r^k\|_{D_{m+n}}^2 - \|\hat{p} - q^k\|_D^2$. Also, using again the results of Lemma* 1*, we get $\|\hat{p} - q^k\|_D^2 = \|y^j - y^0\|_D^2$.*

4.2. Proof of Theorem 1

Proof. To prove (i), let us consider the computation of the distance $\|p^k - \bar{q}^*\|_D^2$ and $\|p^{k+1} - \bar{q}^*\|_D^2$. The first, $\|p^k - \bar{q}^*\|_D^2$, using the intermediate point $q^{k+1} = [x^{k+1}; 0]$, coincides with $\|p^k - q^{k+1}\|_D^2 + \|q^{k+1} - \bar{q}^*\|_D^2 + 2(q^{k+1} - p^k)^T D(\bar{q}^* - q^{k+1})$. Then, using the coordinates of the points represented by p^k, q^{k+1} and $\bar{q}^*$, we obtain $\|p^k - \bar{q}^*\|_D^2 = \|r^k\|_{D_{m+n}}^2 + \|x^{k+1} - x^k\|^2 + \|x^{k+1} - \bar{x}^*\|^2 + 2(x^{k+1} - x^k)^T(\bar{x}^* - x^{k+1})$.

Because $p^{k+1} = [x^{k+1}; r^{k+1}]$, it follows from the approximation $\hat{p} = [z^j; \mu^j]$ satisfying condition (13), we know that $x^{k+1} = z^j$, $r^{k+1} = \mu^j + s^j$, where $s^j = \tilde{A}x^{k+1} - \mu^j - \tilde{b}$.

Furthermore, as a consequence of the iterative process of ACCIM we also know, according to the results in subsection 4.1, that $x^{k+1} - x^k = \tilde{A}^T D_{m+n}(-\mu^j)$. Then, $(x^{k+1} - x^k)^T(\bar{x}^* - x^{k+1}) =$

$-\mu^{j\,T} D_{m+n}(r^* - r^{k+1})$. Taking into account the ACCIM properties (i), (ii) and (iii) of Lemma 1, and its specific application in Remark 2 to the problem solved by Algorithm 2, we obtain $-\mu^{j\,T} D_{m+n}(r^* - r^{k+1}) = -\mu^{j\,T} D_{m+n} r^* + \|\mu^j\|^2_{D_{m+n}}$.

Hence, $\|p^k - \bar{q}^*\|^2_D$ is equal to $\|r^k\|^2_{D_{m+n}} + \|x^{k+1} - x^k\|^2 + \|x^{k+1} - \bar{x}^*\|^2 + 2\|\mu^j\|^2_{D_{m+n}} - 2\mu^{j} D_{m+n} r^*$.

Let us now consider $\|p^{k+1} - \bar{q}^*\|^2_D = \|x^{k+1} - \bar{x}^*\|^2 + \|r^{k+1}\|^2_{D_{m+n}}$. By (b) of Remark 3 we know that $s^{j\,T} D_{m+n} \mu^j = 0$, and $\|r^{k+1}\|^2_{D_{m+n}} = \|s_j\|^2_{D_{m+n}} + \|\mu^j\|^2_{D_{m+n}}$. Therefore, the difference $\|p^k - \bar{q}^*\|^2_D - \|p^{k+1} - \bar{q}^*\|^2_D$ can be described by means $\|r^k\|^2_{D_{m+n}} + \|x^{k+1} - x^k\|^2 + (\|\mu^j\|^2_{D_{m+n}} - 2\,\mu^{j\,T} D_{m+n} r^* + \|r^*\|^2_{D_{m+n}}) - \|r^*\|^2_{D_{m+n}} - \|s^j\|^2_{D_{m+n}}$, after having added and subtracted $\|r^*\|^2_{D_{m+n}}$. Then, considering that $\|\mu^j\|^2_{D_{m+n}} - 2\mu^{j\,T} D_{m+n} r^* + \|r^*\|^2_{D_{m+n}} = \|\mu^j - r^*\|^2_{D_{m+n}} \geqslant 0$, we obtain $\|p^k - \bar{q}^*\|^2_D - \|p^{k+1} - \bar{q}^*\|^2_D \geqslant \|r^k\|^2_{D_{m+n}} + \|x^{k+1} - x^k\|^2 - \|s^j\|^2_{D_{m+n}} - \|r^*\|^2_{D_{m+n}}$.

On the other hand, from (a) of Remark 3 we know that $\|r^k\|^2_{D_{m+n}} = \|q^k - \hat{p}\|^2_D + \|p^k - \hat{p}\|^2_D$. Also, it is known by condition (12) that $\|s^j\|^2_{D_{m+n}} \leqslant \gamma \|p^k - \hat{p}\|^2_D$, with $0 < \gamma \leqslant 1/2$. Then, the difference $\|p^k - \bar{q}^*\|^2_D - \|p^{k+1} - \bar{q}^*\|^2_D \geqslant \|q^k - \hat{p}\|^2_D + (\frac{1}{\gamma} - 1)\|s^j\|^2_{D_{m+n}} - \|r^*\|^2_{D_{m+n}}$. Furthermore, considering that $\|q^k - \hat{p}\|^2_D = \|x^{k+1} - x^k\|^2 + \|\mu^j\|^2_{D_{m+n}}$, and $(\frac{1}{\gamma} - 1)\|s^j\|^2_{D_{m+n}} \geqslant \|s^j\|^2_{D_{m+n}}$ by hypothesis, and $\|r^{k+1}\|^2_{D_{m+n}} = \|\mu^j\|^2_{D_{m+n}} + \|s^j\|^2_{D_{m+n}}$, we obtain that the difference $\|p^k - q^*\|^2_D - \|p^{k+1} - \bar{q}^*\|^2_D \geqslant \|r^{k+1}\|^2_{D_{m+n}} - \|r^*\|^2_{D_{m+n}}$. Therefore, the sequence $\{\|p^k - \bar{q}^*\|^2_D\}$ is decreasing, bounded, and therefore convergent.

Because we know that the sequence $\{\|p^k - \bar{q}^*\|^2_D\}$ converges, we get a similar result for the sequence $\{\|p^k - \bar{p}^*\|^2_D\}$ considering that $\|p^k - p^*\|^2_D = \|(p^k - \bar{q}^*) - (\bar{p}^* - \bar{q}^*)\|^2_D$. Using the orthogonality between $p^k - \bar{p}^*$ and $\bar{p}^* - \bar{q}^*$, and considering that $\bar{p}^* - \bar{q}^* = [0; r^*]$, we get $\|p^k - p^*\|^2_D = \|(p^k - q^*)\|^2_D - \|r^*\|^2_{D_{m+n}}$. Therefore, because the sequence $\{\|p^k - \bar{q}^*\|^2_D\}$ is decreasing and convergent, the sequence $\{\|p^k - \bar{p}^*\|^2_D\}$ also converges.

From the fact that the distances $\{\|p^k - p^*\|^2_D\}$ decrease, $\bar{p}^*$ being the element defined in (19), it follows that the sequence $\{p^k\}$ is in a compact set B_0, centered in $\bar{p}^*$, because it satisfies $\|p^k - \bar{p}^*\|^2_D < \|p^0 - \bar{p}^*\|^2_D$. Hence, a subsequence $\{p^{k_s}\} = \{[x^{k_s}; r^{k_s}]\}$ exists, satisfying $r^{k_s} = \tilde{A} x^{k_s} - \tilde{b}$, convergent to $[\bar{x}; \bar{r}] \in B_0$.

By (iv) of Lemma 3 in Scolnik et al. (2008), we know that $\tilde{A}^T D_{m+n} r^{k_s}$ tends to zero; then, $\tilde{A}^T D_{m+n} \bar{r} = 0$. Therefore, $[\bar{x}; \bar{r}]$ satisfies the optimality condition of problem (15). Moreover, $\bar{r} = r^*$ because of the unicity of the minimal residual r^*. Furthermore, because $\|\bar{x} - \bar{x}^*\| \leqslant \|x^{k_s} - \bar{x}\| + \|x^{k_s} - \bar{x}^*\|$, due to (v) of Lemma 3 in Scolnik et al. (2008), we know that $\|x^{k_s} - \bar{x}^*\| \leqslant \|x^{k_s} - \bar{x}\|$; therefore, $\|\bar{x} - \bar{x}^*\| \leqslant 2\|x^{k_s} - \bar{x}\|$. Hence, we obtain that $\bar{x} = \bar{x}^*$.

Finally, because the sequence of the norms $\|p^k - \bar{p}^*\|_D$ converges, and $\{\|p^{k_s} - \bar{p}^*\|_D\}$ goes to zero, the sequence $\{\|p^k - \bar{p}^*\|_D\}$ tends to zero. Therefore, the sequence $\{p^k\}$ converges to $\bar{p}^*$.

5. Numerical experiments

The main purpose of the experiments reported here is to compare the RIOP performance, applied to the solution of regularized least-squares problems (5), with the one of IOP (1). The latter is

taken as a reference because in Scolnik et al. (2008), we have reported comparisons of IOP with ART (underrelaxed), CAV (Censor et al., 2001), LANDW (Landweber, 1951), and KE (Popa, 1998), showing that IOP was the most efficient algorithm for the test problems and algorithms used for the comparison.

Also in Scolnik et al. (2006), we compared the IOP algorithm with KERP (Popa and Zdunek, 2004), both converging to the minimum norm solution in the case of rank-deficient problems. Those experiments were performed using two image models, "phantoms", kindly provided by Popa and Zdunek, and reported in their paper (Popa and Zdunek, 2005). Although the results were promising, it was possible to see that the quality of the reconstructed images deteriorated after a certain number of iterations. Therefore, the aim of the following experiments is to test the efficiency of the regularized model combined with RIOP.

The algorithms were implemented sequentially, and the experiments were run on a Pentium IV, with 512 MB RAM. In the implementation of IOP and RIOP we consider the parameter $\gamma = 10^{-2}$ in the initial iteration; then, $\gamma = 10^{-1}$.

Both algorithms stop when the condition $|\|r^{k+1}\| - \|r^k\|| < \varepsilon \max(\|r^0\|, 1)$ is satisfied.

For every problem the initial point was $x^0 = 0$.

In order to create the symmetric matrix $M \in \Re^{n\times n}$, we have considered for each pixel $i = 1, \ldots, n$, the weight of its interaction with its neighbors j, $j \in S_i$ by means of $w_{ij} = 1$ if j interacts horizontally or vertically, and $w_{ij} = 1/\sqrt{2}$ for the diagonal interactions. Other possibilities were tested but the above led to the best results. From these assignments it follows that the matrix M has diagonal entries $M_{ii} = \sum_{j\in S_i} w_{ij}$, and $M_{ij} = -1$ if the index j corresponds to a vertical or a horizontal neighbor, and $M_{ij} = -1/\sqrt{2}$ if j indicates a diagonal neighbor.

5.1. Test problems

The experiments were performed using two image models, "phantoms", provided by Popa and Zdunek, reported in their paper (Popa and Zdunek, 2005). They lead to rank-deficient problems and are those used in Scolnik et al. (2006). Also aiming at testing problems with larger dimensions, we considered two cases (B8 and B7), which have characteristics different from the first two, and are given in the SNARK system (Browne et al., 1993), a software package for testing and evaluating algorithms for image reconstruction problems.

The first two simulate real objects in electromagnetic geotomography, whose data come from projections made with a limited angular range (Popa and Zdunek, 2004, 2005). These problems are modeled by means of a system $Ax = b$, where b_i is the attenuation of the electromagnetic field along the i-th ray, and each a_{ij} element of A represents the contribution of the j-th pixel, in relation to the i-th ray, to the attenuation b_i. These problems lead to inconsistent systems, and the corresponding matrix has a deficient rank due to the angle limitations of the projections. The first model (A_1 matrix), whose original image is given in Fig. 1 (left), is associated with an area of 12 m × 12 m in a square of 12 × 12 pixels, the total number of rays being 144. The second model (A_2 matrix), whose original image is presented in Fig. 1 (right), has an area of 30 m × 30 m, represented in a square of 30 × 30 pixels, using 900 rays. The characteristics of matrices A_1 and A_2 are given in Table 1.

We analyze the results obtained with IOP and RIOP with the system $Ax = b + \delta b$, arising from simulating noisy perturbations of the right-hand side b. These two test problems analyzed here are

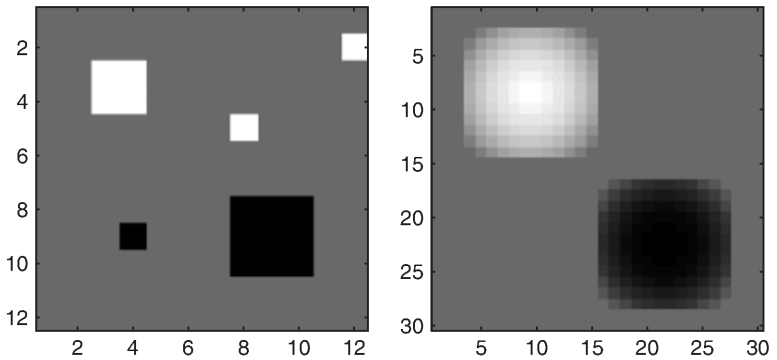

Fig. 1. Original image: A_1 (left), A_2 (right).

Table 1
Properties of A_1 and A_2

Properties	A_1	A_2
$m \times n$	144×144	900×900
Rank	125	805
Cond (A)	$9.39E04$	$2.15E07$
Sparsity	90.1%	95%
x^{exact} (image)	12×12 (pixels)	30×30 (pixels)

exactly those of Popa and Zdunek (2005). Starting from the knowledge of $Ax^{\text{exact}} = b$, a perturbation δb is defined satisfying $\|\delta b\|/\|b\| \approx 5.5\%$. Because $\delta b = \delta b_A + \delta b_{A^\perp}$, where $\delta b_A \in R(A)$ and $\delta b_{A^\perp} \in R(A)^\perp$, the case considered satisfies $\|\delta b_A\| = \|\delta b_{A^\perp}\|$. These perturbations are applied to each problem according to:

$$A_1: \ \|\delta b_{A_1}\| = \|\delta b_{A_1{}^\perp}\| = 1.76,$$
$$A_2: \ \|\delta b_{A_2}\| = \|\delta b_{A_2{}^\perp}\| = 11.3076.$$

The other two problems used in these numerical experiments are from the SNARK system that was only used for generating the test problems B7 and B8. The visualizations of the reconstructed images were obtained with MATLAB 5.3.

As a consequence of the discretization of the image reconstruction problem by X-ray transmission, a Cartesian grid of image square elements called pixels is defined in the region of interest in such a way that it covers the totality of the image to be reconstructed. With a suitable selection of the variables that define the geometry of the data such as angles between rays, the distances between rays when they are parallel, and the localization of the sources, the generated problems have full rank matrices. The systems $Ax = b$ (see Censor et al. (2001) for a more complete description) are generally overdetermined and inconsistent, because in the fully discretized model each line integral of the attenuation along the i-th ray is approximated by a finite sum.

We show the performance of the algorithms on the reconstruction of the Herman head phantom (B8) defined by a set of ellipses, with a specific attenuation value attached to each original elliptical region. This test is one of the same appearing in Censor et al. (2001) and Scolnik

et al. (2008) that arise from considering a different number of projections and number of rays per projection, leading in such a way to systems with a different number of variables and equations. A 115×115 digitization of the phantom is shown in Fig. 2 (right).

Problem B7 simulates the collection of positron emission tomography (PET) data. The test phantom was obtained from a computerized overlay atlas based on the average anatomy of the brain (Herman and Odhner, 1991). In this phantom, neuro-anatomical structures are represented by various ellipses and rectangles at suitable locations. A 95×95 digitization of this phantom is shown in Fig. 2 (left). To simulate data obtained from a PET system with a ring of 300 detectors and with each detector in coincidence with 101 detectors opposite to it, divergent projection data have been generated over 300 view angles with 101 rays per view. In this case it is assumed that the detectors are arranged on an arc. Poisson noise is introduced into the projection measurements and the phantom densities are scaled by a factor of 0.51 (see Browne et al., 1993).

In Table 2 the dimensions of the test problems are shown using m for the number of rows, and n for the number of columns.

5.2. Parameters

The matrix D_m used for problems A_1, A_2, and B8 in the implementation of algorithms was $D_m = diag(\|a_i\|^2)$, that is the one that gave the best performance with respect to the descent of the norm of the residual, among several alternatives. For solving the problem B7 from SNARK, the matrix D_m used in the implementation was $D_m = diag(10^{-2})$, because the residuals of the true

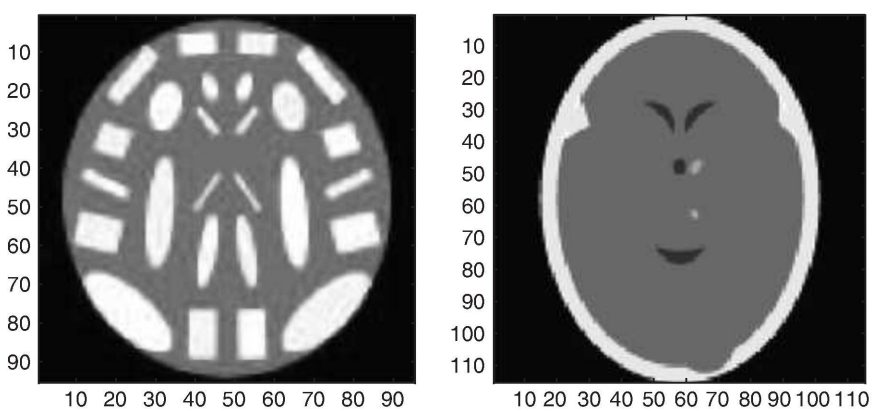

Fig. 2. Original image: B7 (left), B8 (right).

Table 2
Test of SNARK

	Equations (m)	Variables (n)	Image size ($N \times N$)	Projections	Rays
B7	27,376	9025	95×95	300	101
B8	22,303	13,225	115×115	151	175

solution are very large due to the simulation made with the system, and thus it is not convenient to force the residuals to become too small.

The L-curve (Hansen and O'Leary, 1993) has been used to determine the regularization parameter for all problems. In general, the L-curve is a plot of the M-norm of the regularized solution versus the norm of the corresponding residual. In Tikhonov's standard regularization, it exhibits a corner behavior as a function of the regularization parameter β. The regularization parameter is often chosen to be on the corner of the L-curve (Hansen, 1998), because in this region the regularized solution yields both a small residual norm and a small solution M-norm. In our case, in which we use a weighted norm of the residual and the matrix M, the L-shape differs from the classical one although values are detected such that there is an equilibrium between the decrease of the norm of the residual and the decrease of the M-norm of the regularized solution.

We show in Fig. 3 as an example the shape of the L-curve for A_1 and B7, using 50 values of α in the interval $[10^{-4}, 10]$. In order to compare the obtained values, we report the curves of the distance as a function of α, taking into account the definition of the distance to the exact solution as given in Censor et al. (2001). It can be seen that for A_1 (left), acceptable values of the regularization parameter are those of $\alpha \in [0.03, 0.06]$, for B7 (right), $\alpha \in [0.3, 0.6]$. In our implementations, the values used were determined by inspection of the curve, choosing for A_1, $\alpha = 0.05$, for B7, $\alpha = 0.5$. Similarly we obtained $\alpha = 1.2$ for A_2, and $\alpha = 0.8$ for B8.

5.3. Results

With the purpose of displaying the quality of the reconstructed image using IOP with those obtained with RIOP, we include graphs that compare the performance of algorithms, and curves representing the distance of the obtained density with regard to the one of the true images (Censor et al., 2001), and also the quality of the reconstructed image by each algorithm.

Distance: between the solution x^k (k-th iteration) and the true image x^{exact}, which is computed by means of $\dfrac{\left\|x^k - x^{\text{exact}}\right\|}{\sqrt{N}\sigma_{\text{exact}}}$, with $\sigma_{\text{exact}} = \dfrac{\left\|x^{\text{exact}} - \rho_{\text{exact}}\right\|}{\sqrt{N}}$, where $\rho_{\text{exact}} = \dfrac{\sum_{j=1}^{N} |x_j^{\text{exact}}|}{N}$.

In Fig. 4 (left), we compare the performances of IOP and RIOP by means of the distance of the reconstructed images for problem A_1. In Fig. 5, the reconstructed images corresponding to A_1 are shown for IOP and RIOP, respectively. It is necessary to point out that the number of iterations are the internal ones for IOP and RIOP. The criterion used for reporting the obtained images at the specified iterations was to choose those closer to 150–250–500, as given in Popa and Zdunek (2005).

In Fig. 4 (right) the performance of IOP and RIOP is compared in relation to the distance for problem A_2. In Fig. 6, we show for problem A_2 the reconstructed images by IOP and RIOP corresponding to the inner iterations closest to 150, 250, and 500. As can be seen in the curves and images, RIOP reaches the image closest to the original for both test problems A_1 and A_2. RIOP also obtains the minimum distance, and keeps it below the one corresponding to IOP, between the 50th and the 500th iterations. It also obtains the best reconstructions in that interval, obtaining the best image in the neighborhood of the 250th iteration.

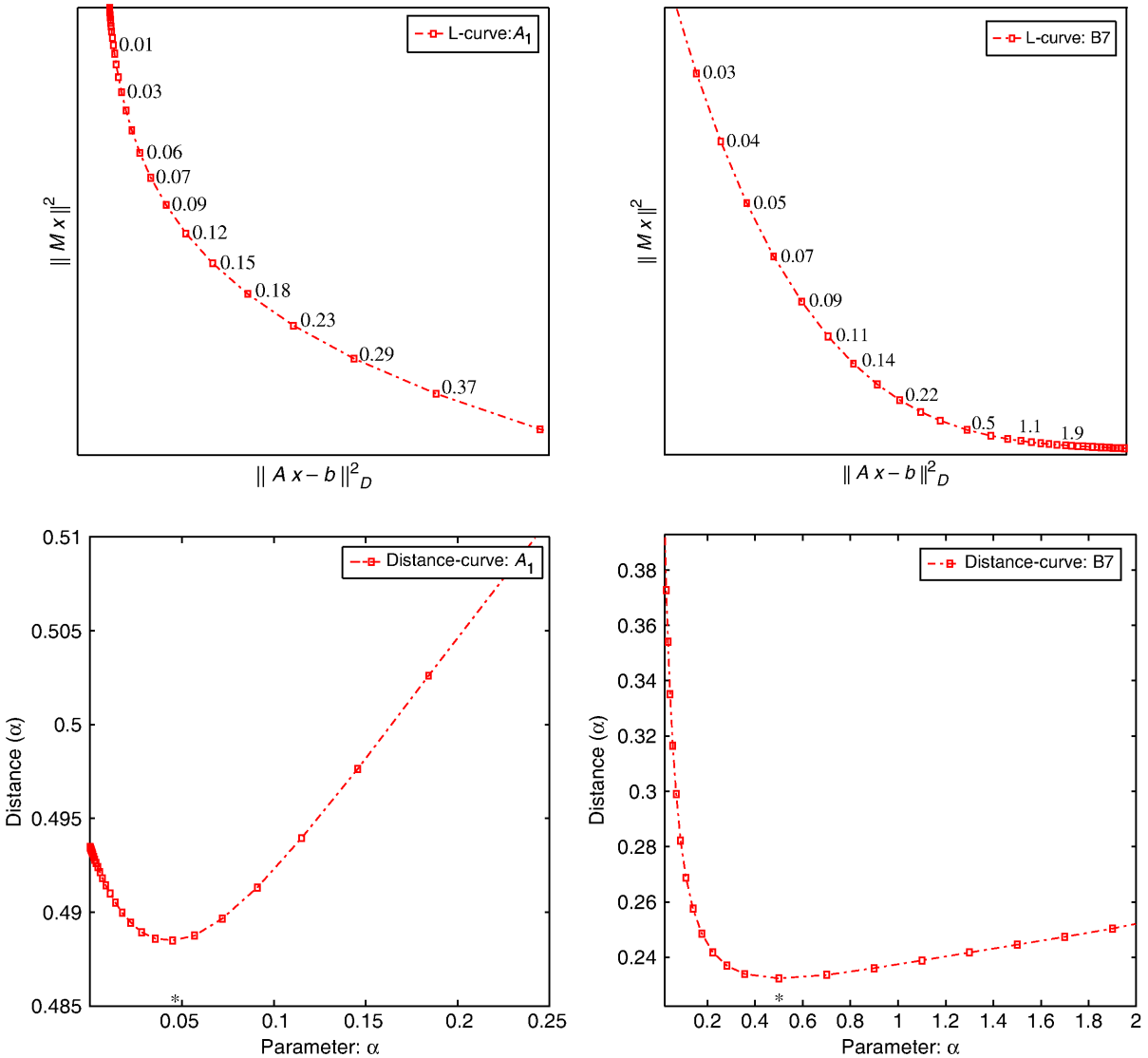

Fig. 3. L-curve (top), distance-curve (bottom).

In the following, the results obtained for problems B7 and B8 are shown. In Fig. 7 (left), the performance of IOP and RIOP is compared by means of the distance for problem B7. In Fig. 8, we show for the same problem the reconstructed images by both algorithms corresponding to the inner iterations closest to 30, 40, and 60.

In Fig. 7 (right), the performance of IOP and RIOP is compared in relation to the distance for problem B8, while in Fig. 9 we show for the same problem, the reconstructed images by both algorithms corresponding to the inner iterations closest to 10, 15, and 20. Also, in Fig. 10, we show for the same problem the reconstructed images by those algorithms corresponding to the

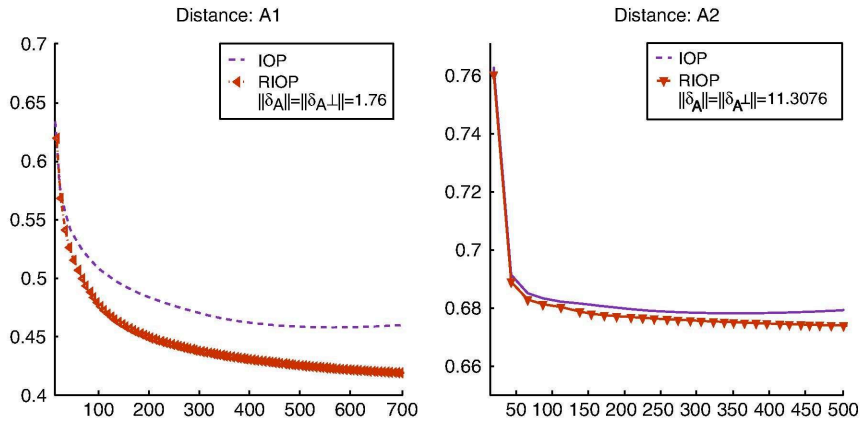

Fig. 4. Distance IOP–RIOP: A_1 (left), A_2 (right).

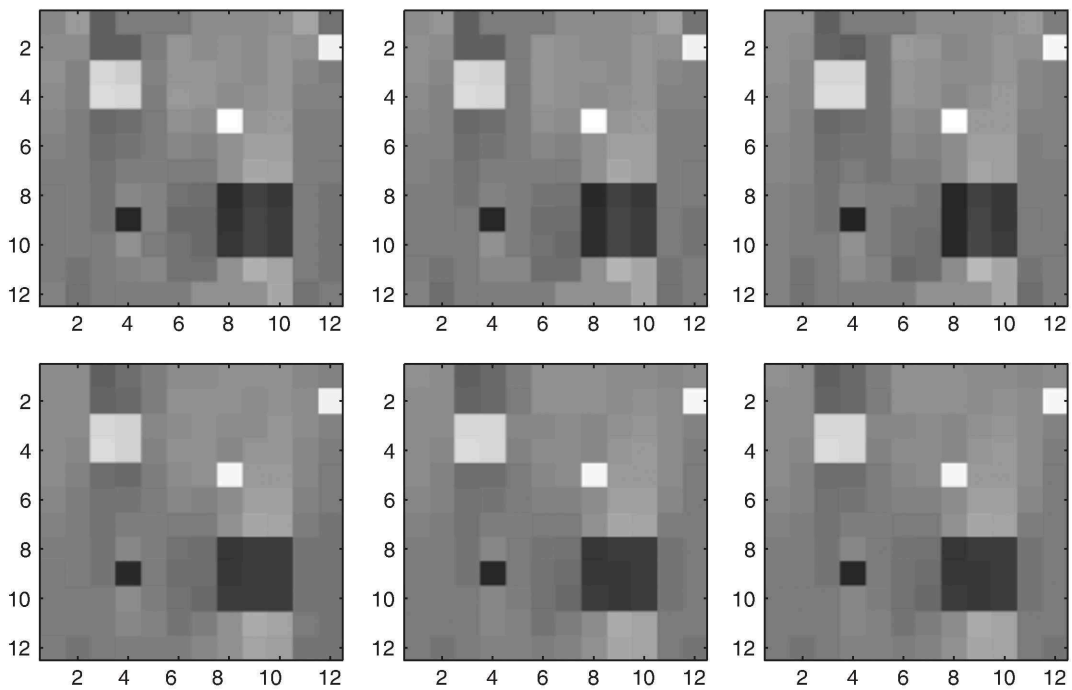

Fig. 5. A_1: IOP (top)–RIOP (bottom) (150–250–500 iterations).

inner iterations closest to 80, aiming at comparing the differences when the number of iterations is increased. In problems B7 and B8, it can be seen in the curves and images that RIOP reaches the image closest to the original for both test problems.

Both IOP and RIOP rapidly decrease the distances in the first iterations, but RIOP continues to decrease and keeps them below those of IOP. It also obtains the best reconstructions from the point of view of roughness. It is very important to point out that by enlarging the images it can be

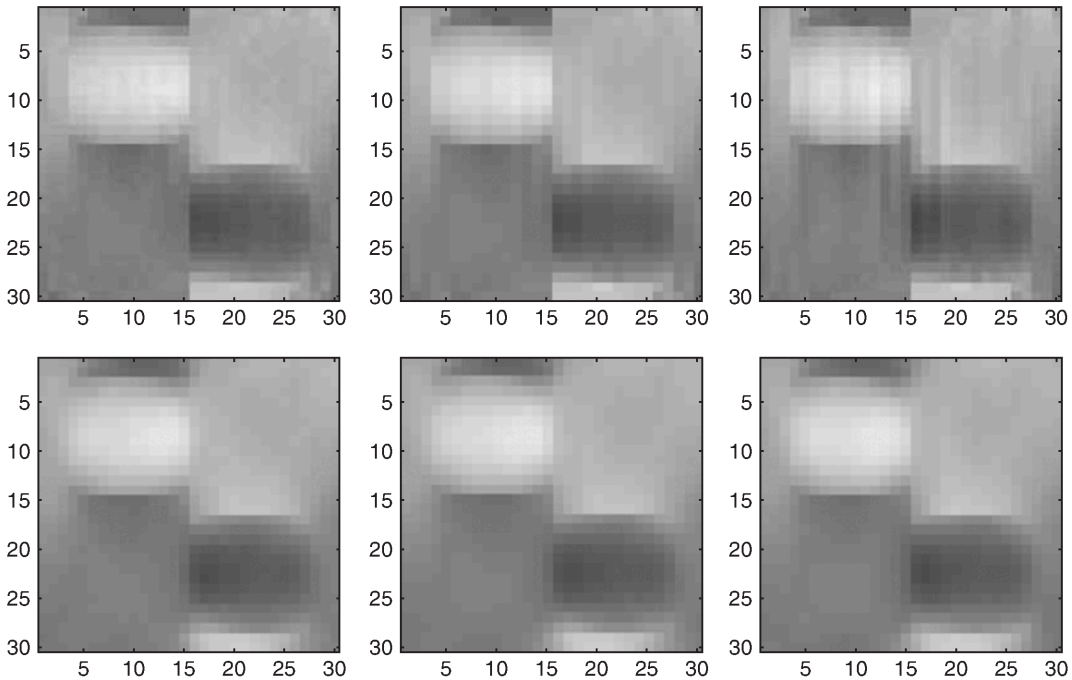

Fig. 6. A_2: IOP (top)–RIOP (bottom) (150–250–500 iterations).

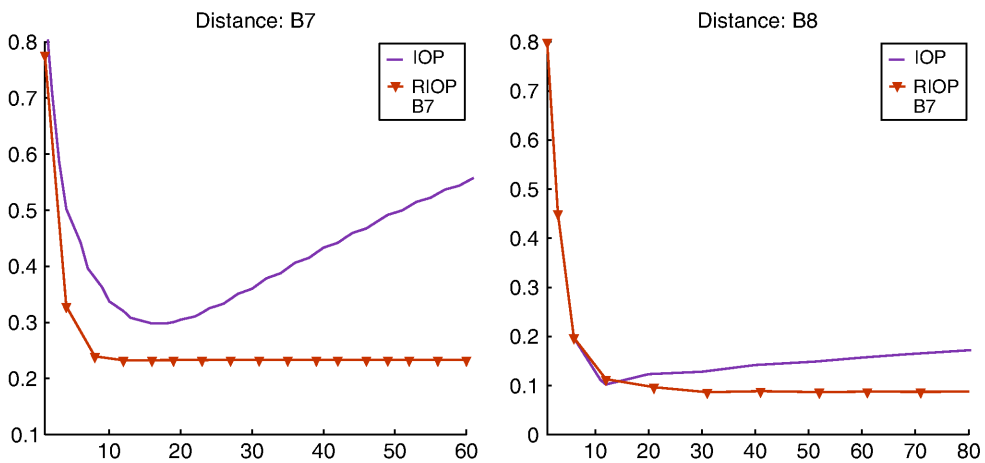

Fig. 7. Distance IOP–RIOP: B7(left), B8(right).

seen that the quality obtained with IOP deteriorates after a certain number of iterations (Fig. 10), while RIOP maintains both the distance and the quality of the reconstructed images. In Table 3, we give the CPU times required by each algorithm for reaching a solution satisfying the convergence criterion with different tolerances (parameter $\varepsilon = 10^{-5}, 5.10^{-5}, 10^{-4}$). We include in the table the obtained distance to the exact image for each case. It is worth noting that RIOP solves an augmented system, although the matrix M of the aggregated system is sparse and with a

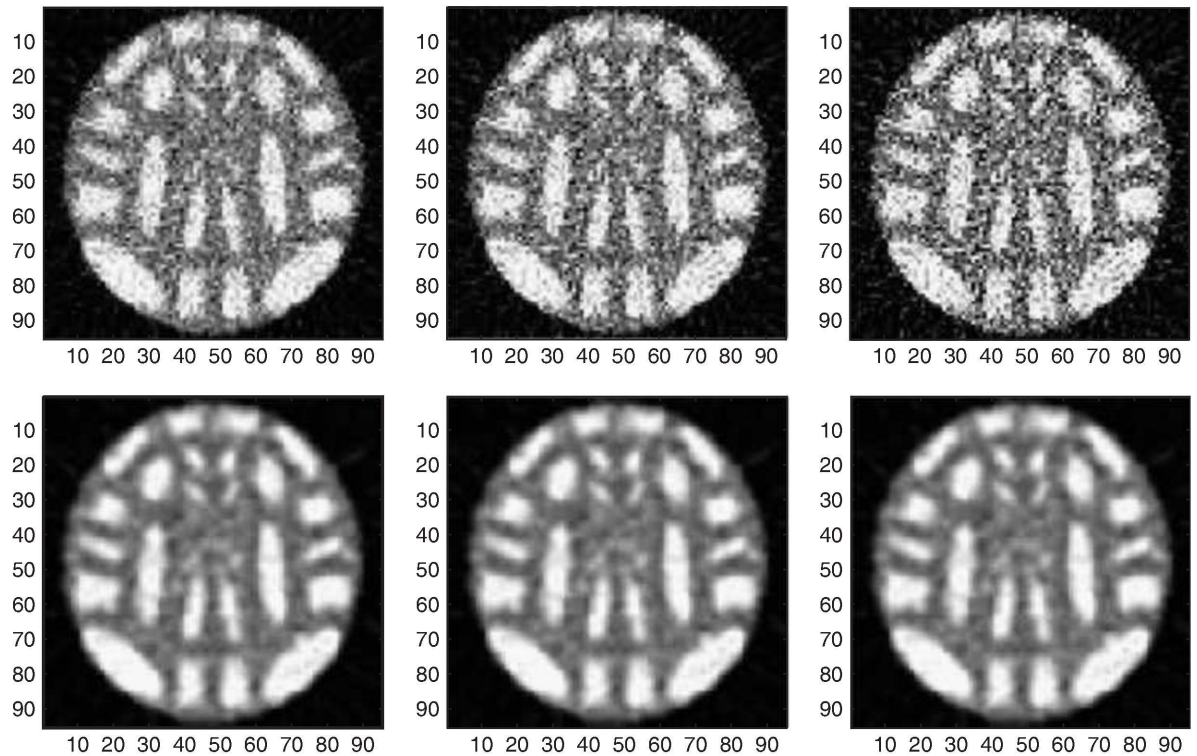

Fig. 8. B7: IOP (top)–RIOP (bottom), 30–40–60 iterations.

known structure (no more than 9 nonzero entries per row). Hence, the CPU time is not increased in a sensible way as can be seen in the table.

6. Conclusions

In the field of ECT (emission computed tomography), iterative reconstruction techniques include the maximum likelihood expectation maximization (ML-EM), ordered subsets expectation maximization (OS-EM), rescaled block iterative expectation maximization (RBI-EM) (Byrne, 1998), and row action maximization likelihood (RAMLA) techniques (Browne and De Pierro, 1996). However, and just to show possible uses of the RIOP approach, we restricted ourselves to the use of what can be considered as similar algorithms, without claiming that the results obtained are better than those that can be given by RAMLA, etc. In fact, to use the framework presented here (combined with, for instance, blobs (Lewitt, 1990)) is something that can be done in the future.

We recall that the aim of this paper was to present a modification of least-squares model (1) used for image reconstruction problems. That was because the minimum norm solution does not always turn out to be the closest to the true image due to the underlying discretized model. Hence,

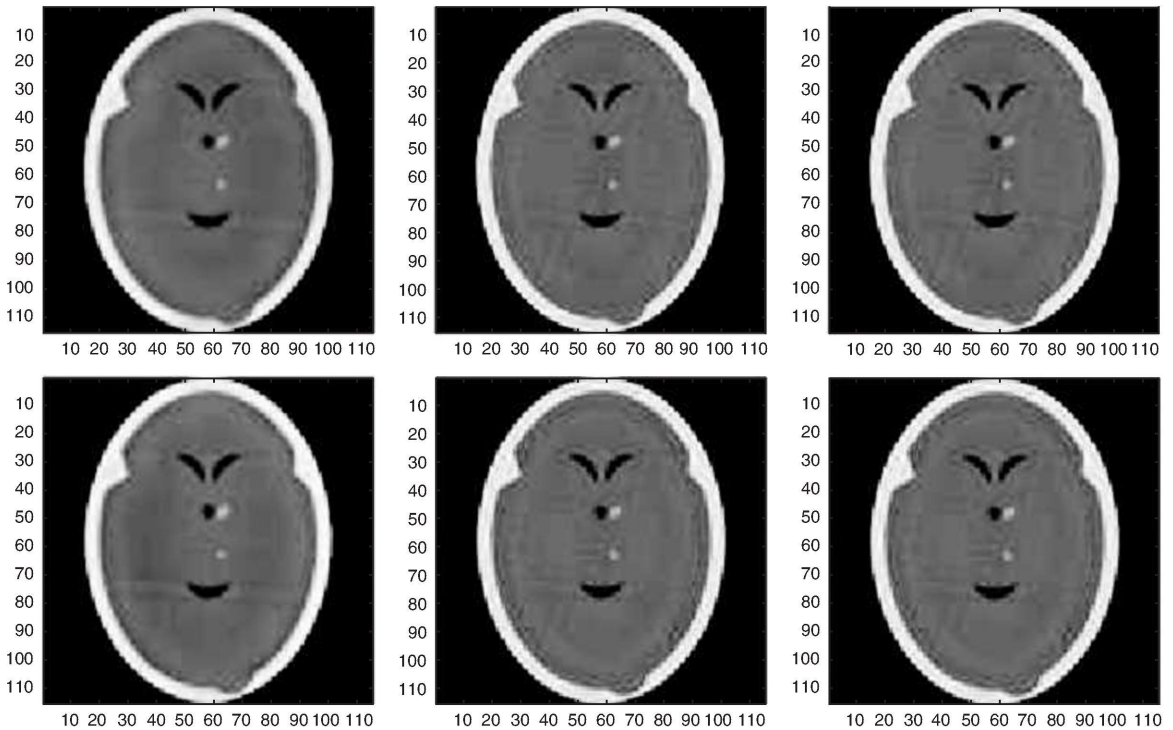

Fig. 9. B8: IOP (top)–RIOP (bottom), 10–15–20.

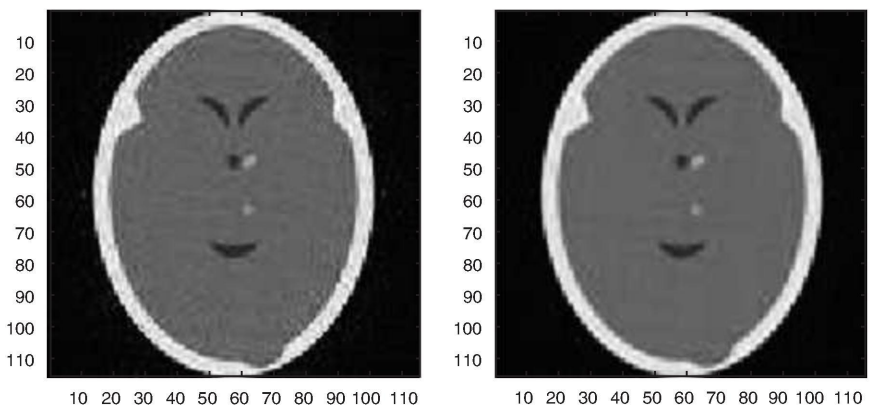

Fig. 10. B8: IOP (left)–RIOP (right), 80 iterations.

we have studied an approach for regularizing model (1) in order to match convergence with image quality. As the numerical results show, the RIOP effectiveness is remarkable in several problems. However, when the inhomogeneity comprises an area larger than the considered neighborhood, the actual regularization is not that effective (case A_2) because it does not reduce the parasite smearings meaningfully.

Table 3
CPU time required for reaching convergence

Problem	Method	Iter	CPU time (s)	Distance
B7	RIOP	56	9.20	0.232445
$\varepsilon = 10^{-5}$	IOP	227	35.42	0.959450
B7	RIOP	38	6.23	0.232704
$\varepsilon = 5.10^{-5}$	IOP	102	15.86	0.718537
B7	RIOP	32	5.31	0.234127
$\varepsilon = 10^{-4}$	IOP	68	10.61	0.587310
B8	RIOP	91	15.27	0.087620
$\varepsilon = 10^{-5}$	IOP	142	22.47	0.207481
B8	RIOP	61	10.31	0.087616
$\varepsilon = 5.10^{-5}$	IOP	83	13.19	0.174564
B8	RIOP	61	10.25	0.087616
$\varepsilon = 10^{-4}$	IOP	64	10.17	0.161811

IOP, incomplete oblique projections; RIOP, regularized incomplete oblique projections.

In forthcoming papers, we will analyse alternative functions for a regularized least-squares problem, aiming at smoothing images. Likewise, for the same $R(x)$ used in this paper, we will study the effect of adding neighboring pixels expanding the radius, and also what suitable weights should be chosen from a practical viewpoint. Another open question is to estimate the regularization parameter, together with the regularization function $U(x)$. It is worthwhile to mention that methods that converge to a minimal-norm solution compute an optimum point in which both regularization errors and perturbation errors are well balanced (Björk, 1996; Hansen and O'Leary, 1993). This point corresponds to the corner on the L-curve (Hansen, 1998) or is the minimun of the GCV criterion (Björk, 1996, p. 212). Also, other authors (Pralat and Zdunek, 2005) state that the parameter β is roughly estimated with the curves of $U(x^*)$ versus β, where $x^* = x^*(\beta)$ is a final solution obtained after a suitable number of iterations at β. The minima of the curves should determine the optimum values of β for each reconstruction. This is because the total energy function $U(x)$ can be interpreted as a measure of the total roughness of the image. For low values of β, the solution is dominated by the classical least-squares model. For high values of β, the regularization term dominates in the solution and the effect of oversmoothing may take place.

Acknowledgements

We would like to thank C. Popa and R. Zdunek for having kindly provided their test problems used in Popa and Zdunek (2004), and to the anonymous referees for their constructive suggestions.

References

Björk, A., 1996. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia.

Browne, J.A., De Pierro, A.R., 1996. A row-action alternative to the EM Algorithm for maximizing likelihoods in emission tomography. *IEEE Transactions on Medical Imaging* 15, 4, 687–699.

Browne, J.A., Herman, G.T., Odhner, D., 1993. SNARK93: a programming system for image reconstruction from projections. Department of Radiology, University of Pennsylvania, Medical Image Processing Group, Technical Report MIPG198.

Byrne, C., 1998. Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative method. *IEEE Transactions on Image Processing* 7, 1, 100–109.

Byrne, C., 2002. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Problems* 18, 441–453.

Byrne, C., Censor, Y., 2001. Proximity function minimization using multiple Bregman projections with applications to split feasibility and Kullback–Leibler distance minimization. *Annals of Operations Research* 105, 77–98.

Censor, Y., Elfving, T., 2002. Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem. *SIAM Journal on Matrix Analysis and Applications* 24, 40–58.

Censor, Y., Gordon, D., Gordon, R., 2001. Component averaging: an efficient iterative parallel algorithm for large and sparse unstructured problems. *Parallel Computing* 27, 777–808.

Censor, Y., Zenios, S., 1997. *Parallel Optimization: Theory and Applications*. Oxford University Press, New York.

Csiszár, I., Tusnády, G., 1984. Information geometry and alternating minimization procedures. *Statistics and Decisions* 1, (Suppl), 205–237.

Echebest, N., Guardarucci, M.T., Scolnik, H.D., Vacchino, M.C., 2005. An accelerated iterative method with diagonally scaled oblique projections for solving linear feasibility problems. *Annals of Operations Research* 138, 235–257.

García Palomares, U.M., 1993. Parallel projected aggregation methods for solving the convex feasibility problem. *SIAM Journal of Optimization* 3, 882–900.

Hansen, P.C., 1998. *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, Philadelphia.

Hansen, P.C., O'Leary, D.P., 1993. The use of L-curve in the regularization of discrete ill-posed problems. *SIAM Journal of Scientific Computing* 14, 1487–1503.

Herman, G.T., Odhner, D., 1991. Perfomance evaluation of an iterative image reconstruction algorithm for positron emission tomography. *IEEE Transactions on Medical Imaging* 10, 336–346.

Jiang, M., Wang, G., 2003. Convergence studies on iterative algorithms for image reconstruction. *IEEE Transactions on Medical Imaging* 22, 569–579.

Kaczmarz, S., 1937. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin of the International Academy of Polonaise Science Letters* 35, 355–357.

Koltracht, I., Lancaster, F., Smith, D., 1990. The structure of some matrices arising in tomography. *Linear Algebra Applications* 130, 193–218.

Landweber, L., 1951. An iteration formula for Fredholm integral equations of the first kind. *American Journal of Mathematics* 73, 615–624.

Lange, K., 1990. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Transactions on Medical Imaging* 9, 439–446.

Lewitt, R.M., 1990. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. *Journal of the Optical Society of America A* 7, 10, 1834–1846.

Luenberger, D.G., 1986. *Linear and Nonlinear Programming*. Addison Wesley Publishing Company, New York.

Popa, C., 1998. Extensions of block-projections methods with relaxation parameters to inconsistent and rank-deficient least-squares problems. *BIT* 38, 151–176.

Popa, C., Zdunek, R., 2004. Kaczmarz extended algorithm for tomographic image reconstruction from limited data. *Mathematics and Computers in Simulation* 65, 579–598.

Popa, C., Zdunek, R., 2005. Penalized least squares image reconstruction for borehole tomography. *Proceedings of Algorithmy 2005*, 260–269.

Pralat, A., Zdunek, R., 2005. Electromagnetic geotomography – selection of measuring frequency. *IEEE Sensors Journal* 5, 2, 242–250.

Scolnik, H.D., Echebest, N., Guardarucci, M.T., Vacchino, M.C., 2002a. A class of optimized row projection methods for solving large non-symmetric linear systems. *Applied Numerical Mathematics* 41, 499–513.

Scolnik, H.D., Echebest, N., Guardarucci, M.T., Vacchino, M.C., 2002b. Acceleration scheme for parallel projected aggregation methods for solving large linear systems. *Annals of Operations Research* 117, 95–115.

Scolnik, H.D., Echebest, N., Guardarucci, M.T., 2006. Extensions of incomplete oblique projections method for solving rank-deficient least-squares problems. Proceedings of Latin-Ibero American Congress of Operations Research, Montevideo, Uruguay, November 27–30 (submitted to the *Journal of Industrial and Management Optimization*, 2007).

Scolnik, H.D., Echebest, N., Guardarucci, M.T., Vacchino, M.C., 2008. Incomplete oblique projections for solving large inconsistent linear systems. *Mathematical Programming B* 111, 273–300.

Tikhonov, A.N., 1963. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady* 4, 1035–1038.

Wahba, G., 1977. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal of Numerical Analytics* 14, 651–667.

Wahba, G., 1990. Spline models for observational data. DBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, PA.