Proceedings of ALGORITMY 2005
pp. 260–269

PENALIZED LEAST-SQUARES IMAGE RECONSTRUCTION FOR BOREHOLE TOMOGRAPHY

CONSTANTIN POPA* AND RAFAL ZDUNEK†

Abstract. The Algebraic Reconstruction Technique (ART), which is based on the Kaczmarz's method, is very popular in many tomographic applications of image reconstruction. However, this algorithm gives satisfactory approximations to the exact solution only for consistent data. For inconsistent problems (which is a real case in practice), the reconstruction is speckled with noise, and the convergence is not asymptotical. In a previous paper we made a systematic analysis, both theoretical and experimental, for this case by using an extension of the classical Kaczmarz's algorithm. But, although the results were much better and very promising comparing them with some classical and widely used projection algorithms, another difficulty still could not be eliminate, namely the exact solution of the discrete problem is not always enough accurate. This aspect is of course related to the mathematical modeling of the real problem (image), which gives us only a classical least-squares formulation. In this paper, we considered a penalized least-squares objective. The penalty term is defined with the Gibbs prior that incorporates nearest neighbor interactions among adjacent pixels to enforce an overall smoothness in the image. Then, we derived a version of the above mentioned extended Kaczmarz algorithm for this new penalized least-squares problem. Our synthetic experiments showed that this algorithm has not only a good noisy performance, but it also moderates parasite effects of the limited angular ray-coverage in borehole tomography. The effects are visible in the form of vertical smearings from inhomogeneous features in the image.

Key words. linear least-squares problems, extended Kaczmarz algorithm, regularized image reconstruction, borehole tomography

AMS subject classifications. 65F10, 65F20, 65F22, 86A22

1. The extended Kaczmarz algorithm. Many problems in the field of tomographic image reconstruction are modeled by the (classical) linear least-squares problem: find $x^* \in \mathbb{R}^n$ such that

$$\| Ax^* - b \| = \min!, \tag{1.1}$$

where A is an $m \times n$ real matrix and $b \in \mathbb{R}^m$ a given vector (by $\| \cdot \|$ and $\langle \cdot, \cdot \rangle$ we shall denote the Euclidean norm and scalar product on some space $\mathbb{R}^q$). Let $LSS(A, b)$ be the set of all least-squares solutions of (1.1) and x_{LS} its minimal norm one. Also $B^t, R(B), N(B)$ will denote the transpose, range and null space of a matrix B, respectively, and P_S the orthogonal projection onto a closed convex set $S \subset \mathbb{R}^q$. By $a_i \in \mathbb{R}^n, \alpha_j \in \mathbb{R}^m$, we shall denote the i-th row and j-th column of A, respectively and we shall suppose that

$$a_i \neq 0, \ \forall i = 1, \ldots, m, \alpha_j \neq 0, j = 1, \ldots, m. \tag{1.2}$$

Moreover, all the vectors that will appear will be considered as column vectors. We have the equality (see e.g. [3])

$$LSS(A; b) = x_{LS} + N(A). \tag{1.3}$$

*Faculty of Mathematics and Computer Science, "Ovidius" University, Blvd. Mamaia 124, 900527, Constanta, Romania (cpopa@univ-ovidius.ro); this author was supported by the PNCDI INFOSOC Grant 131/2004

†Institute of Telecommunications and Acoustics, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland (Rafal.Zdunek@pwr.wroc.pl)

Moreover, with respect to the decomposition of b as

$$b = b_A + b_A^*,\ b_A = P_{R(A)}(b),\ b_A^* = P_{N(A^t)}(b) \tag{1.4}$$

we have the equivalent formulations of (1.1)

$$x^* \in LSS(A;b) \Leftrightarrow Ax^* = b_A, \Leftrightarrow A^tAx^* = A^tb. \tag{1.5}$$

Problems of the form (1.1), coming from the field of tomographic image reconstruction are in general sparse, without full rank (rank-defficient) and inconsistent (because of measurements related to the right-hand side b). Concerning the first property - sparsity - the fill-in process will make direct methods (as e.g. QR decomposition) prohibitively costly in terms of storage and operations (see e.g. the comments and examples in [3], pag. 270). On the other hand, the rank-defficiency will not allow us to use well known efficient iterative solvers as LSQR (see e.g. the above cited paper, pag. 307). And, with respect to the last above mentioned "bad" property, beside their computational appeal (simplicity, row-action nature, etc) and their efficacious performance in some specific situations (see e.g. [4]), some algorithms from the ART class methods have been successfuly extended for the inconsistent case of (1.1). In this respect we shall consider in our paper the Kaczmarz extended algorithm with relaxation parameters, firstly proposed in [15] and successfuly used for image reconstruction in borehole tomography in [16] and [17]. It will be briefly described in what follows. For this let $i = 1, \ldots, m$ and the applications

$$f_i(b;x) = x - \frac{\langle x, a_i\rangle - b_i}{\| a_i \|^2} a_i,\ F(b;x) = (f_1 \circ \cdots \circ f_m)(b;x), \tag{1.6}$$

and the classical Kaczmarz algorithm (for short **K**), widely used in ART.
Algorithm (K): Let $x^0 \in I\!R^n$; for $k = 0, 1, 2, \ldots$ do

$$x^{k+1} = F(b;x^k). \tag{1.7}$$

The following convergence result was proved in [14] (see also [15]).

THEOREM 1.1. *(i) The sequence $(x^k)_{k\geq 0}$ from (1.7) always converges and*

$$\lim_{k\to\infty} x^k = P_{N(A)}(x^0) + Gb, \tag{1.8}$$

where the $n \times n$ matrix G is a generalized inverse of A.
(ii) If the problem (1.1) is consistent, i.e.

$$b \in R(A) \Leftrightarrow b = b_A, \tag{1.9}$$

then the limit in (1.8) is a solution of (1.1). Moreover, we have

$$x_{LS} = Gb. \tag{1.10}$$

The first problem that appears in real world applications is that due to modeling and measurement errors the formulation (1.1) is usually inconsistent, i.e.

$$b \notin R(A) \Leftrightarrow b_A^* \neq 0, \tag{1.11}$$

In this case, the limit in (1.8) given by the classical Kaczmarz algorithm (1.6) is no more one of its least-squares solutions. More clear, if we denote by $x^*(x^0)$ the limit in (1.8), in [16], we proved the following result:

THEOREM 1.2. *For any limit point $x^*(x^0)$ in (1.8) we have the equality*

$$d\left(x^*(x^0), LSS(A;b)\right) \geq \| Gb_A^* \|, \tag{1.12}$$

where d denotes the Euclidean distance.

One possibility for overcoming this first problem is to use an extension of the classical Kaczmarz algorithm, proposed by one of the authors in [15]. We shall briefly describe it in what follows in a generalized version which uses relaxation parameters. For this, let $\alpha \neq 0, \omega \neq 0$ be two real numbers the applications (see for details [15])

$$f_i(\omega; b; x) = (1-\omega)x + \omega f_i(b; x),\ F(\omega; b; x) = (f_1 \circ \cdots \circ f_m)(\omega; b; x), \tag{1.13}$$

with $f_i(b;x)$ from (1.6) and

$$\varphi_j(y) = y - \frac{\langle y, \alpha_j \rangle}{\| \alpha_j \|^2}\alpha_j,\ \Phi(y) = (\varphi_1 \circ \cdots \circ \varphi_n)(y), \tag{1.14}$$

$$\varphi_j(\alpha; y) = (1-\alpha)y + \alpha\varphi_j(y),\ \Phi(\alpha; y) = (\varphi_1 \circ \cdots \circ \varphi_n)(\alpha; y). \tag{1.15}$$

With all these elements we can define the **K**aczmarz **E**xtended with **R**elaxation **P**arameters algorithm as follows.

Algorithm KERP. Let $x^0 \in I\!R^n, y^0 = b$; for $k = 0, 1 \ldots$ do

$$y^{k+1} = \Phi(\alpha; y^k), \tag{1.16}$$

$$b^{k+1} = b - y^{k+1}, \tag{1.17}$$

$$x^{k+1} = F(\omega; b^{k+1}; x^k). \tag{1.18}$$

The following convergence result was proved in [15].

THEOREM 1.3. *For any $x^0 \in I\!R^n$, any $\omega, \alpha \in (0,2)$ and any right hand side b in (1.1), the sequence $(x^k)_{k\geq 0}$ generated by the above algorithm* **KERP** *converges always to one of the least-squares solutions of (1.1).*

Unfortunately, there is still an unpleasant problem concerned with the numerical solution of (1.1). It is independent on its consistency/inconsistency and is related to the fact that the minimal norm solution x_{LS} (which is usually calculated) is not always an enough good approximation of the exact solution (image). In this sense we can refer at the situation presented in Fig. 4.2 (see [16] for details). The theoretical explanation for this starts from the characterization (1.3). According to this, let $x^* \in LSS(A;b)$ be the exact image (see Fig. 4.1). From (1.3) and the fact that x_{LS} is orthogonal to $N(A)$ (see e.g. [16]) we get

$$x^* - x_{LS} = P_{N(A)}(x^*). \tag{1.19}$$

Thus, if the component on $N(A)$ of the exact image x^* is "enough big", we can expect that x_{LS} is not an accurate approximation of it. On the other hand, the matrix A depends only on the pixels and rays that are generated during the mathematical modeling of the real problem (see for details [16]), thus there can exist images with such a big $N(A)$ component (and this is exactly the situation in Fig. 4.2). Then one of the ways to overcome this second difficulty is to reformulate the mathematical modeling of our problem. In this sense we proposed the penalized least-squares formulation which is described in details in the next section.

2. The regularized Kaczmarz algorithm. Since in our application the problem (1.1) is very ill-posed, we can consider a more general problem which is the regularized weighted least-squares problem: find $x^* \in R^n$

$$||Ax^* - b||^2_{W^{-1}} + \beta R(x^*) = \min! \tag{2.1}$$

where W is a matrix which attributes weights to data, β is a regularization parameter, and $R(x)$ is a measure of roughness in the image. Usually, $R(x)$ is defined by a square of discrete smoothing norm $R(x) = ||Lx||^2$, where L is typically either an identity or diagonal matrix [8] (p12). In our consideration, we wish that $R(x)$ would provide not only the information about total roughness in the image but also the prior information about the missing image components. Both kinds of the information can be easily incorporated to the image under a form of the following Gibbs prior that is commonly used in statistical image reconstruction [5]:

$$\pi(x) = \frac{\exp\{-\beta U(x)\}}{\int \exp\{-\beta U(x)\} dx}. \tag{2.2}$$

Total energy function $U(x)$ in (2.2) is usually defined as

$$U(x) = \sum_{j=1}^{N} \sum_{n \in S_j} w_{j,n} V(x_j - x_n, \delta) \tag{2.3}$$

where S_j is a set of indices of the nearest neighborhood of pixel j, $w_{j,n}$ is weighting factor defined by the Markov Random Field (MRF), and $V(x_j - x_n, \delta)$ is a potential function.

In statistical image reconstruction the maximization problem is a subject to be solved

$$\max_x p(x|b) \tag{2.4}$$

where $p(x|b)$ is a posterior given by the Bayes' theorem

$$p(x|b) = \frac{p(b|x)\pi(x)}{\int p(b|x)\pi(x) dx}.$$

The solution to (2.4) can be also found solving the minimization problem

$$\min_x \Psi(x), \quad \text{where} \quad \Psi(x) = -2 \ln p(x|b) \tag{2.5}$$

Assuming perturbation of data have the nature of an additive zero-mean Gaussian noise, the relation between data and an expectation of data can be modelled by a Gaussian statistics, i.e.

$$p(b|x) = (2\pi)^{-m/2} |W|^{-1/2} \exp\{-\frac{1}{2}(b - Ax)^t W^{-1} (b - Ax)\},$$

hence

$$\Psi(x) = ||Ax - b||^2_{W^{-1}} + 2\beta U(x) + const \tag{2.6}$$

Thus, the discrete smoothing norm in (2.1) can be defined as $R(x) = 2U(x)$.

To solve problem (2.5) a simultaneous iterative scheme can be applied, which can be done with a generalized Landweber scheme, or a preconditioned gradient based scheme [10]. Convergence properties of the latter were also discussed in [13]. Thus

$$x^{(k+1)} = x^{(k)} - \lambda_k M^{-1} \nabla \Psi(x^{(k)}) \tag{2.7}$$

where λ_k is a relaxation factor, and M is a preconditioner. In our case

$$\nabla \Psi(x) = A^t W^{-1}(Ax - b) + 2\beta \nabla U(x).$$

We can also apply a sequential block-iterative scheme, which leads to the following algorithm

$$x^{(k+1)} = x^{(k)} + \lambda_k M^{-1} \left[A_{[k]}^t W_{[k]}^{-1} \left(b_{[k]} - A_{[k]} x^{(k)} \right) - 2\beta \nabla U(x^{(k)}) \right]. \tag{2.8}$$

where $[k] = k(\mod K) + 1$, and K is a total number of blocks into which A was divided. Assuming $K = m$, $A_{[k]}$ is the $[k]$-th row of A where $[k] = k(\mod m) + 1$. The same rule concerns $W_{[k]}^{-1}$ and $b_{[k]}$. Moreover, assuming $M = I$ and

$$W^{-1} = \operatorname{diag}\left(||a_i||^2 \right), \quad i = 1, \ldots, M$$

where a_i is the i-th row of A, algorithm (2.8) boils down to the *regularized Kaczmarz's algorithm with relaxation parameter*. Thus, the regularized KERP (**RKERP**) algorithm can be written as

Algorithm RKERP: Let $x^{(0)} \in I\!R^N$, $y^{(0)} = b$, for $k = 0, 1, \ldots$, do

$$y^{(k+1)} = \Phi(\alpha; y^{(k)}), \tag{2.9}$$

$$b^{(k+1)} = b - y^{(k+1)}, \tag{2.10}$$

$$x^{(k+1)} = F(\omega; b^{(k+1)}; x^{(k)}) - 2\beta \nabla U(x^{(k)}), \tag{2.11}$$

where $F(\omega; b^{(k+1)}; x^{(k)})$ and $\Phi(\alpha; y^{(k)})$ are given by (1.13) and (1.15), respectively. We assume $\lambda_k = 1$ in (2.8) but the relaxation is now introduced through ω in (2.11).

The potential functions in (2.3) attributes, so called, "clique energy" to a pair of adjacent pixels. Table 2.1 lists the potential functions that are the most popular in the literature on image processing. Since the Green's function satisfies all the properties mentioned in [12], i.e. it is nonnegative, even, 0 at $r = 0$, strictly increasing for $r > 0$, unbounded, convex, and has bounded first-derivative, we decided to select this function to our tests. Thus $\nabla U(x^{(k)})$ in (2.11) takes the form

$$\frac{\partial}{\partial x_j} U(x)\big|_{x=x^{(k)}} = \sum_{n \in S_j} w_{j,n} \tanh\left(\frac{x_j - x_n}{\delta} \right), \quad j = 1, \ldots, N \tag{2.12}$$

3. Numerical experiments. The tests were carried on for the data (image A and B) used in [16], and additionally, for the data generated for a new original image – image C illustrated in Fig. 4.1 (right). Each reconstruction is performed from noisy data for which $SNR = 25dB$, and $||\delta b_A|| = ||\delta b_A^*||$. Hence, we set the following levels of perturbation $||\delta b_A|| = ||\delta b_A^*|| = 1.76$, $||\delta b_A|| = ||\delta b_A^*|| = 11.3076$ and $||\delta b_A|| = ||\delta b_A^*|| = 7.5513$ for images A, B and C, respectively. The setting of parameters α, β, δ and ω for each case are given in Table 3.1. Parameter $w_{j,n}$ in (2.3) should be set according to the MRF, i.e. $w_{j,n} = 1$ for horizontal and vertical interaction lines, and $w_{j,n} = \frac{1}{\sqrt{2}}$ for diagonal interaction lines. Additionally, image B was also reconstructed for the modified MRF for which $w_{j,n} = \frac{1}{3}, 3, \frac{1}{\sqrt{(1/3)^2 + 3^2}}$ for horizontal, vertical, and diagonal interaction lines, respectively.

TABLE 2.1
Potential functions

.

Author(s) (Name)	Reference	Functions: $V(r,\delta)$
(Gaussian)		$\left(\frac{r}{\delta}\right)^2$
Besag (Laplacian)	[2]	$\left\|\frac{r}{\delta}\right\|$
Hebert and Leahy	[9]	$\delta \log\left[1+(\frac{r}{\delta})^2\right]$
Geman and McClure	[6]	$\frac{16}{3\sqrt{3}}\frac{(r/\delta)^2}{(1+(r/\delta)^2)}$
Geman and Reynolds	[7]	$\frac{\lvert r/\delta\rvert}{1+\lvert r/\delta\rvert}$
Stevenson and Delp (Hubert)	[18]	$\min\{\lvert\frac{r}{\delta}\rvert^2, 2\lvert\frac{r}{\delta}\rvert - 1\}$
Green	[5]	$\delta \log[\cosh(r/\delta)]$

TABLE 3.1
Parameters of reconstruction

.

	α	β	δ	ω
Image A	0.15	5×10^{-4}	2×10^{-3}	1.4
Image B	6.5×10^{-2}	5×10^{-4}	2×10^{-3}	0.6
Image C	6.5×10^{-2}	2.5×10^{-3}	2×10^{-3}	0.6

4. Conclusions and future work. Comparing the results obtained with the **KERP** and **RKERP**, we can conclude that this kind of regularization is very efficient if the image to be reconstructed contains small (with reference to the resolution), possibly sharp inhomogeneities. This takes place in image A and C. Moreover, Fig. 4.4 shows that the **RKERP** algorithm makes possible to reconstruct the so-called "ghosts" (vertical smearings from inhomogeneities) [11], which means that it recovers some missing image components from $N(A)$. Unfortunately, if the inhomogeneities are much bigger than the nearest neighborhood of a pixel (a square of 9 pixels), which happens in image B, the "correction" information in total energy function (2.3) is too poor to reduce the parasite smearings. Probably, this would work better if we consider a bigger neighborhood (the third and fourth lines of interaction). Note that modification of the MRF noticeably improved the reconstruction (Figs. 4.8 and 4.9).

The reconstruction of image C was performed for the same values of parameters α and ω as for image B. Hence, we can expect that the results in Figs. 4.10 and 4.11 could be better.

The drawback of the **RKERP** algorithm is a necessity of estimating four parameters α, β, δ and ω before reconstruction. Hopefully, the underlying problem of the reconstruction is based on the Bayesian rule, and therefore, we could estimate these parameters through maximization of the Type II likelihood. The problems related with estimation of the parameters, and a selection of a more general prior, such as the inhomogeneous Gaussian random field [1], will be a subject of our future work.

Acknowledgment. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] R. G Aykroyd, S. Zimeras, *Inhomogeneous Prior Models for Image Reconstruction*, J. Am. Stat. Assoc., 94 (1999), pp. 934–946.

[2] J. Besag, *Toward Bayesian Image Analysis*, J. Appl. Stat., 16 (1989), pp. 395–407.

[3] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM Philadelphia, 1996.

[4] Y. Censor, S.A. Zenios , *Parallel optimization. Theory, algorithms, and applications*, Oxford University Press, 1997.

[5] P. J. Green, *Bayesian Reconstruction from Emission Tomography Data Using a Modified EM algorithm*, IEEE Trans. Medical Imaging, 9 (1990), pp. 84–93.

[6] S. Geman, D. McClure, *Statistical Methods for Tomographic Image Reconstruction*, Bull. Int. Stat. Inst., LII-4 (1987), pp. 5–21.

[7] S. Geman, G. Reynolds, *Constrained Parameters and the recovery of Discontinuities*, IEEE Trans. Pattern Anal. Machine Intell., 14 (1992), pp. 367–383.

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

[9] T. Hebert, R. Leahy, *A Generalized EM Algorithm for 3-D Bayesian Reconstruction form Poisson Data Using Gibbs Priors*, IEEE Trans. Medical Imaging, 8 (1989), pp. 194–202.

[10] M. Jiang, G. Wang, *Convergence Studies on Iterative Algorithms for Image Reconstruction*, IEEE Trans. Medical Imaging, 22 (2003), pp. 569–579.

[11] I. Koltracht, P. Lancaster, D. Smith, *The Structure of Some Matrices Arising in Tomography*, Linear Algebra and Its Applications, 130 (1990), pp. 193–218.

[12] K. Lange, *Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing*, IEEE Trans. Medical Imaging, 9 (1990), pp. 439–446.

[13] S. F. McCormick, G. H. Rodrigue, *A Uniform Approach to Gradient Methods for Linear Operator Equations*, J. Math. Analy. Applicat. 49 (1975), pp. 275–285.

[14] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley and Sons, New York, 1986.

[15] C. Popa, *Extensions of Block-Projections Methods with Relaxation Parameters to Inconsistent and Rank-Deficient Least-Squares Problems*; *B I T*, 38(1) (1998), pp. 151-176.

[16] C. Popa, R. Zdunek, *Kaczmarz Extended Algorithm for Tomographic Image Reconstruction from Limited-Data*; Math. and Computers in Simulation, 65 (2004), pp. 579-598.

[17] C. Popa, R. Zdunek, *New generalized oblique projections in DW algorithm with application to borehole tomography*; Proceedings of The Third Workshop on Mathematical Modelling of Environmental and Life Sciences Problems, 27-30 May 2004, Constanta, Romania; Editura Academiei Romane, Bucuresti 2004; 233-243.

[18] R. Stevenson, E. Delp, *Fitting Curves with Discontinuities* in Proc. 1st Int. Workshop on Robust Computer Vision, Oct. 1–3, 1990, pp. 127–126,

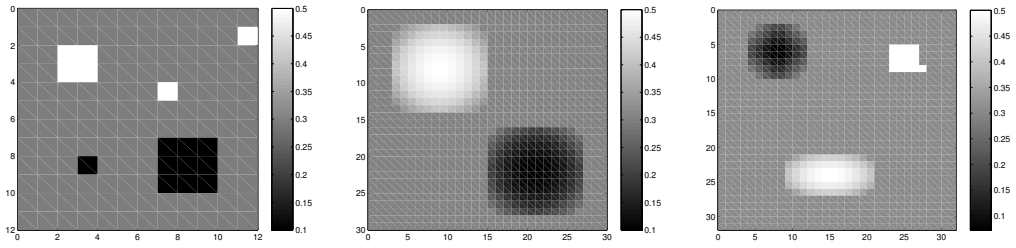

Fig. 4.1. *Original images A, B and C, respectively*

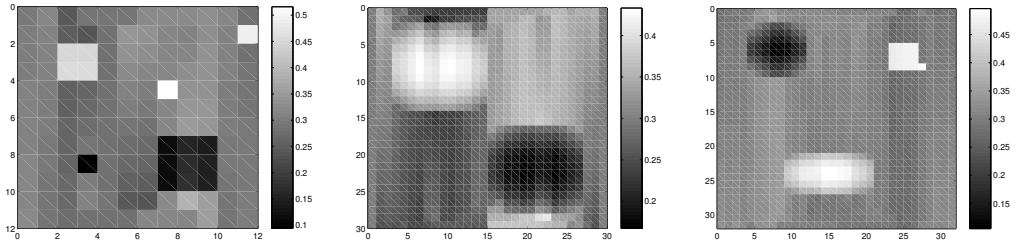

FIG. 4.2. *Minimal-norm least squares solutions (x_{LS}) for images A, B and C, respectively*

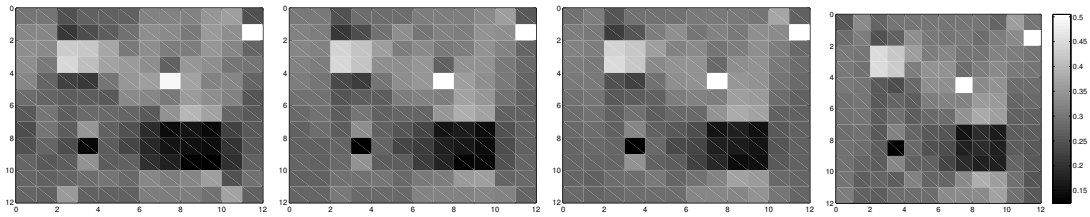

FIG. 4.3. *Image reconstruction with* **KERP** *for $k = 50, 150, 250, 500$; Image A; $||\delta b_A|| = ||\delta b_A^*|| = 1.76$*

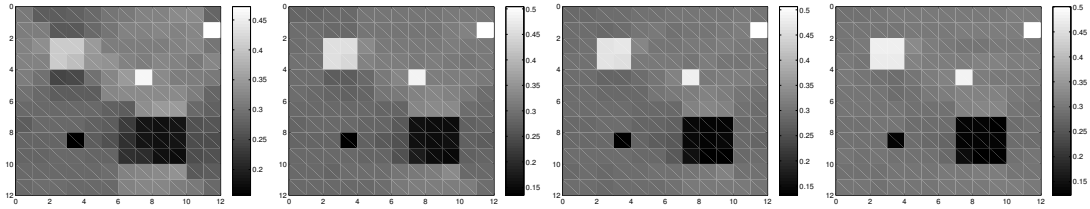

FIG. 4.4. *Image reconstruction with* **RKERP** *for $k = 50, 150, 250, 500$; Image A; $||\delta b_A|| = ||\delta b_A^*|| = 1.76$*

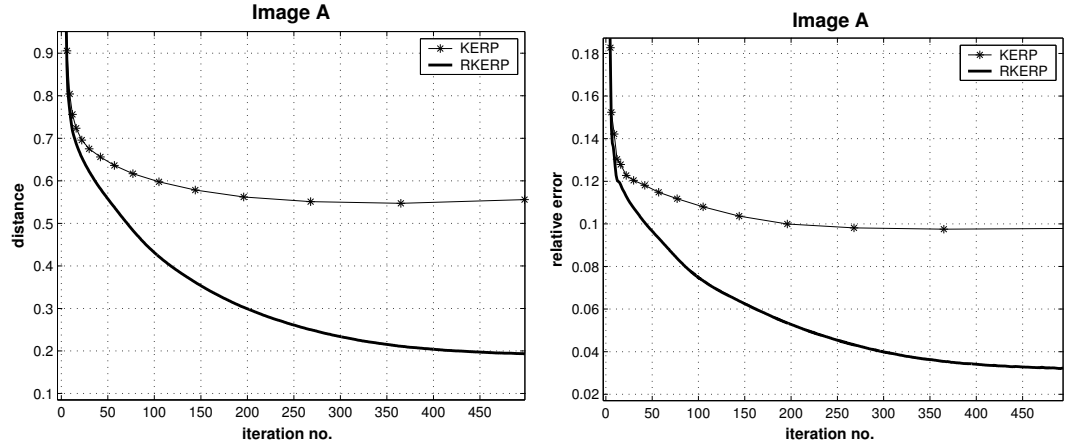

FIG. 4.5. *Distance (left) and relative errors (right) between $x^{(k)}$ and original image x_{exact} for image A*

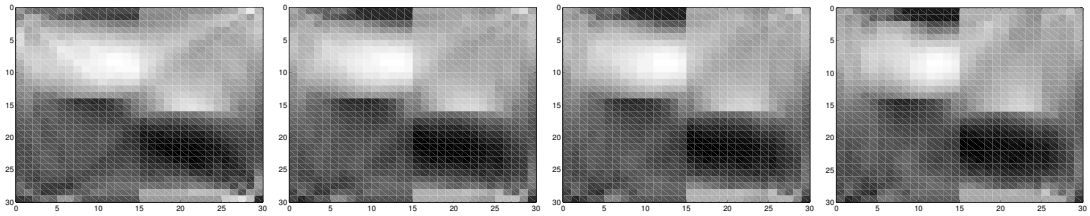

FIG. 4.6. *Image reconstruction with* **KERP** *for* $k = 50, 150, 250, 500$*; Image B;* $||\delta b_A|| = ||\delta b_A^*|| = 11.3076$

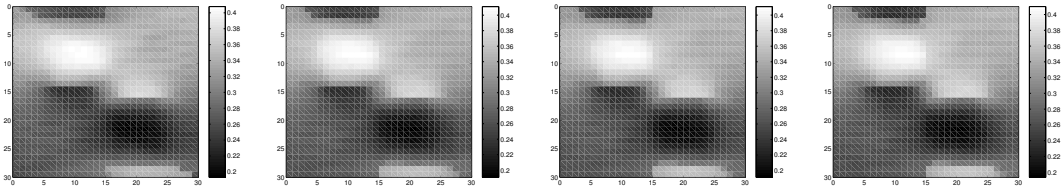

FIG. 4.7. *Image reconstruction with* **RKERP** *for* $k = 50, 150, 250, 500$*; Image B;* $||\delta b_A|| = ||\delta b_A^*|| = 11.3076$*; Normal MRF*

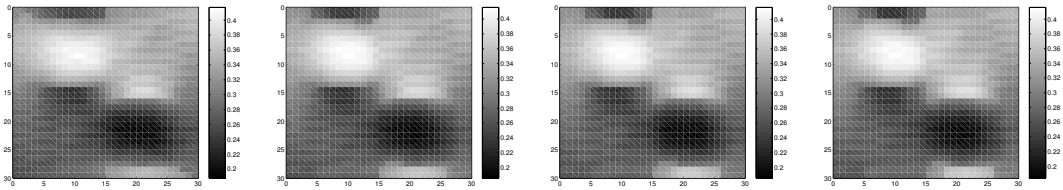

FIG. 4.8. *Image reconstruction with* **RKERP** *for* $k = 50, 150, 250, 500$*; Image B;* $||\delta b_A|| = ||\delta b_A^*|| = 11.3076$*; Modified MRF*

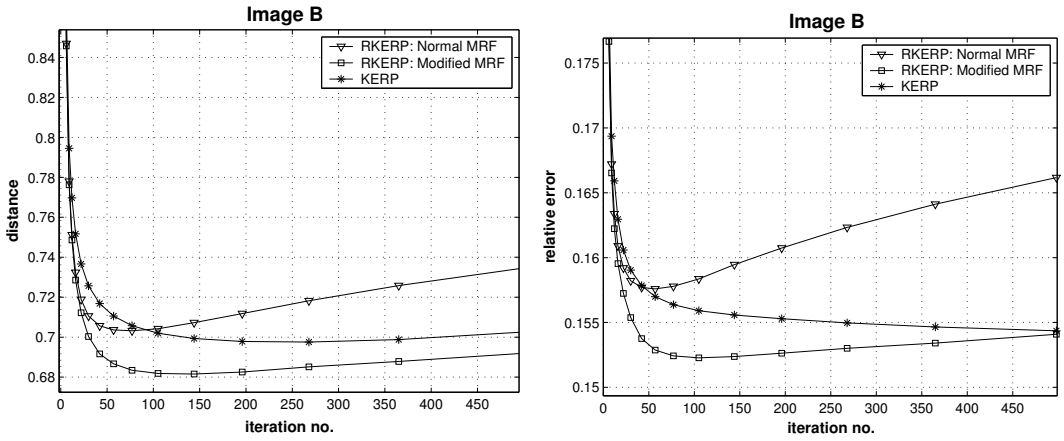

FIG. 4.9. *Distance (left) and relative errors (right) between* $x^{(k)}$ *and original image* x_{exact} *for image B*

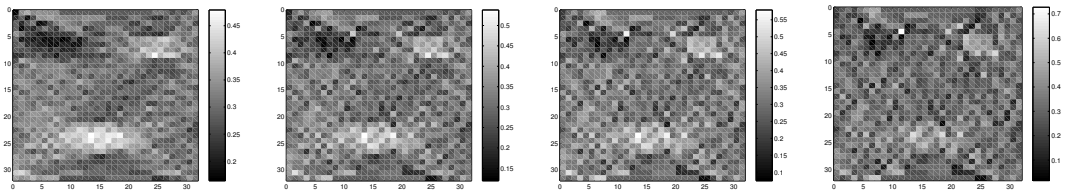

FIG. 4.10. *Image reconstruction with* **KERP** *for* $k = 50, 150, 250, 500$*; Image C;* $||\delta b_A|| = ||\delta b_A^*|| = 7.5513$

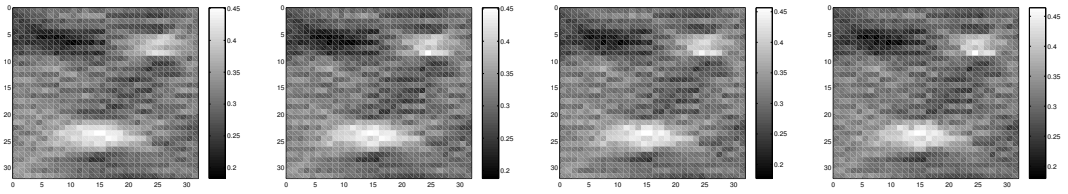

FIG. 4.11. *Image reconstruction with* **RKERP** *for* $k = 50, 150, 250, 500$*; Image C;* $||\delta b_A|| = ||\delta b_A^*|| = 7.5513$

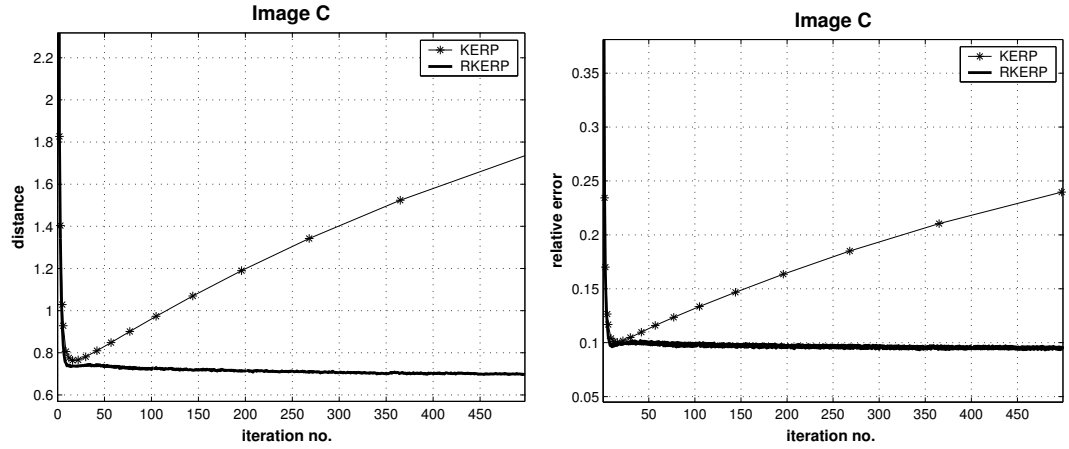

FIG. 4.12. *Distance (left) and relative errors (right) between* $x^{(k)}$ *and original image* x_{exact} *for image C*