Statistica Sinica **5**(1995), 77-96

THE ITERATIVE IMAGE SPACE RECONSTRUCTION ALGORITHM (ISRA) AS AN ALTERNATIVE TO THE EM ALGORITHM FOR SOLVING POSITIVE LINEAR INVERSE PROBLEMS

G. E. B. Archer and D. M. Titterington

University of Glasgow

Abstract. The paper reviews the iterative Image Space Reconstruction Algorithm (ISRA) for solving Linear Inverse Problems with Positive Constraints. The development follows that for the EM algorithm in Vardi and Lee (1993). The algorithm is set down, a range of special cases for particular contexts are listed, convergence issues are discussed, and there is a concluding discussion. The speeds of convergence of EM and ISRA are comparable, although the latter often needs noticeably fewer operations per iteration.

Key words and phrases: EM algorithm, emission tomography, image analysis, image space reconstruction algorithm, integral equations, inverse problems, linear equations, mixtures, motion blurring, portfolio optimization.

1. Introduction to the Iterative Image Space Restoration Algorithm

The iterative Image Space Restoration Algorithm (ISRA) was christened by Daube-Witherspoon and Muehllehner (1986) in the context of image reconstruction from emission computed tomography data. Suppose, in that context, there are M source pixels, the ith of which has emission density f_i. Measurements $\{g_j\}$ are observed, where g_j is the number of coincidences counted in the jth of N pairs of detector elements. There are numbers $\{h_{ij}\}$ such that h_{ij} is the probability that an event emitted from pixel i is received at detector j. Thus $\sum_j h_{ij} = 1$ and the expected value of g_j is $\sum_i f_i h_{ij}$. The reconstruction problem is that of inferring values for the $\{f_i\}$, given the $\{g_j\}$ and assuming that the $\{h_{ij}\}$ are known. The iterative procedure derived by Daube-Witherspoon and Muehllehner (1986) for reconstructing the $f = \{f_i\}$ is as follows: initialize the procedure by choosing an $f^{(o)} > 0$ (i.e. $f_i^{(o)} > 0$ for all i), and then generate a sequence $\{f^{(n)}\}$ using the iteration

$$f_i^{(n)} = f_i^{(n-1)}\Big(\sum_{j=1}^{N} h_{ij} g_j\Big) \Big/ \Big\{\sum_{j=1}^{N} h_{ij}\Big(\sum_{s=1}^{M} f_s^{(n-1)} h_{sj}\Big)\Big\},\ i = 1, \ldots, M,\ n = 1, 2, \ldots. \tag{1.1}$$

This is the discrete version of the ISRA. Daube-Witherspoon and Muehllehner (1986) formulated the ISRA as an alternative to the Poisson-based EM algorithm, for which the iterative step corresponding to (1.1) is

$$f_i^{(n)} = f_i^{(n-1)} \sum_{j=1}^{N} \Big\{ h_{ij} \Big/ \sum_{s=1}^{M} f_s^{(n-1)} h_{sj} \Big\} g_j,\ i = 1, \ldots, M,\ n = 1, 2, \ldots. \tag{1.2}$$

They favoured the ISRA because it involves fewer calculations per iteration than does EM when the number of detector-pairs, N, is large, although the speeds of convergence of the two algorithms are comparable. It is clear, from (1.1), that the summations over j need be carried out only once, at the beginning of the procedure. Daube-Witherspoon and Muehllehner (1986) offered no formal justification for the algorithm, nor did they provide any theoretical arguments justifying its convergence. They motivated the ISRA heuristically by noting that, on the right-hand side of (1.1), $\sum_j h_{ij} g_j$ represents a back-projection of the data $\{g_j\}$ associated with pixel i, whereas $\sum_j h_{ij} \{\sum_s f_s^{(n-1)} h_{sj}\}$ represents a corresponding back-projection of the current 'fitted values' for the $\{g_j\}$. The ISRA is designed to lead eventually to matching of these two sets of back-projections.

Progress on formal justification and convergence properties was made by Titterington (1987) and De Pierro (1987). Titterington (1987) noted that the algorithm could be interpreted as an iterative approach to the computation of least-squares estimates, and De Pierro (1987) also identified the algorithm with a more general algorithm of Chahine (1970) for solving linear equations.

Before we review the work on this interpretation towards the end of Section 2, we derive more general, discrete and continuous versions of the ISRA appropriate for solving a more general class of LININPOS problems (Linear Inverse problems with Positivity restrictions), which may or may not have probabilistic origins. We use the nomenclature and notation of Vardi and Lee (1993), hereafter referred to as VL, and the main purpose of this paper is to show that it is straightforward to develop the ISRA, along the same lines as VL's work on the EM algorithm. Section 3 demonstrates this for a number of special cases. Section 4 discusses issues of convergence for the ISRA, and Section 5 describes both comparisons with the EM algorithm and further generalizations.

2. The EM and ISRA Algorithms for a Class of LININPOS Problems

There are various forms for LININPOS problems, depending on the nature of the underlying structure, and the following equations summarize two different cases:

$$g_j = \sum_{i=1}^{M} f_i h_{ij}, \quad j = 1, \ldots, N; \tag{2.1}$$

$$g(y) = \int_{D_f} h(x,y) f(x) dx, \quad y \in D_g. \tag{2.2}$$

In (2.2), D_f and D_g are the domains of, respectively, the nonnegative real-valued functions f and g, and h is a non-negative, real-valued, bounded function on $D_f \times D_g$. In the discrete version (2.1), $D_f = \{1, \ldots, M\}, D_g = \{1, \ldots, N\}$ and g_j, f_i and h_{ij} are non-negative for all i and j.

The objective is to invert (2.1) and (2.2); given $g(\cdot)$ in (2.2) to solve for $f(\cdot)$ and, given $\{g_j\}$ in (2.1), to solve for $\{f_i\}$. VL point out that both of (2.1) and (2.2) are special cases of the formulation

$$G(y) = \int_{D_f} F(dx) H(x,y), \quad y \in D_g, \tag{2.3}$$

where $G(\cdot)$ and, for any $x \in D_f$, $H(x, \cdot)$ are non-negative measures on D_g, and $F(\cdot)$ is a non-negative measure on D_f. They propose iterative algorithms, similar to the EM algorithm, for inverting the equations. If $G(\cdot)$ and $H(x, \cdot)$, for any x, are probability measures, then so must be $F(\cdot)$, and in this case the algorithms are, for (2.1) and (2.2), respectively, iteration (1.2), initialized by $f_i^{(o)} > 0$, $i = 1, \ldots, M$, and

$$f^{(n)}(x) = f^{(n-1)}(x) \int_{D_g} \Big\{ h(x,y) \Big/ \int_{D_f} f^{(n-1)}(s) h(s,y) ds \Big\} g(y) dy, \quad x \in D_f, \tag{2.4}$$

initialized by $f^{(o)}(x) > 0$, for all $x \in D_f, n = 0, 1, \ldots$. See Titterington and Rossi (1985) for an earlier account of the discrete case.

The ISRA's for solving (2.1) and (2.2) are, respectively, iteration (1.1), initialized by $f_i^{(o)} > 0, i = 1, \ldots, M$, and

$$\begin{aligned} &f^{(n)}(x) \\ &= f^{(n-1)}(x) \Big\{ \int_{D_g} h(x,y) g(y) dy \Big\} \Big/ \Big[\int_{D_g} h(x,y) \Big\{ \int_{D_f} f^{(n-1)}(s) h(s,y) ds \Big\} dy \Big], \\ &\qquad x \in D_f, \end{aligned} \tag{2.5}$$

initialized by $f^{(o)}(x) > 0$, for all $x \in D_f$. These formulae hold whether or not the measures $G(\cdot)$ and $\{H(x, \cdot) : x \in D_f\}$ in (2.3) are probability measures, whereas the formulae of VL require normalization modifications in order to cover the case of general, non-negative measures.

VL discuss the convergence of (1.2) and (2.4) to a solution of the appropriate version of (2.3), they relate the general structure to problems in statistics and optimal portfolio investment, and they expound and illustrate some special cases based on motion-blurred images.

The formulae of VL are related to solutions of maximum (log)likelihood problems and, consequently, to the minimization of certain Kullback-Leibler directed divergences, an observation which is a key point in convergence proofs. The ISRA, on the other hand, is motivated by its relationship to least-squares estimation, as hinted in Section 1.

To see this consider, first, the discrete case, and the sum of squares function

$$S(f) = \sum_{j=1}^{N}\Big(g_j - \sum_{i=1}^{M} f_i h_{ij}\Big)^2 = \|g - H^T f\|_2^2, \tag{2.6}$$

where $g^T = (g_1, \ldots, g_N)$, $f^T = (f_1, \ldots, f_M)$ and $H = \{h_{ij}\}$. We wish to minimize this, subject to $f \geq 0$. Any minimizer $\hat{f}$ of $S(f)$ satisfies

$$HH^T\hat{f} = Hg,$$

i.e.,

$$\sum_j h_{ij}\Big(\sum_s \hat{f}_s h_{sj}\Big) = \sum_j h_{ij} g_j, \quad i = 1, \ldots, M, \tag{2.7}$$

confirming the desired equality of the two sets of back-projections identified by Daube-Witherspoon and Muehllehner (1986); see Section 1.
Thus

$$1 = \Big(\sum_j h_{ij} g_j\Big) \Big/ \Big\{\sum_j h_{ij}\Big(\sum_s \hat{f}_s h_{sj}\Big)\Big\}, \quad i = 1, \ldots, M.$$

Multiplication of both sides by $\hat{f}_i$ leads to a set of equations that clearly stimulate (1.1).

The sum of squares function underlying (2.5) is

$$S(f) = \int_{D_g} \Big\{g(y) - \int_{D_f} f(s)h(s,y)ds\Big\}^2 dy,$$

and application of the calculus of variations provides the stationarity equations

$$\int_{D_g} h(x,y)\Big\{\int_{D_f} f(s)h(s,y)ds\Big\}dy = \int_{D_g} h(x,y)g(y)dy. \tag{2.8}$$

Iteration (2.5) evolves from (2.8) in the same way that (1.1) does from (2.7).

Clearly, in both the discrete and continuous cases, if the initial $f^{(o)} > 0$, then $f^{(n)} \geq 0$ for all n, in view of the nonnegativity of $G(\cdot)$ and $\{H(x,\cdot)\}$. Even if the latter are all probability measures, there is no guarantee that $f^{(n)}$ is a probability measure for any n, except possibly for $n = 0$, by design, but, if the algorithms converge to a solution of (2.1) or (2.2), then the limit will be a probability measure if the same is true for $G(\cdot)$ and $\{H(x, \cdot)\}$. If only the $\{H(x, \cdot)\}$ are probability measures, then the algorithm converges to a solution with total measure equal to that of G. In principle, one might constrain the total measure associated with F using a Lagrange multiplier (c.f. Section 4.5.2 of Titterington et al. (1985)), but it seems an unnecessary complication in view of the limiting behaviour.

It is important to emphasize that, although both the EM and ISRA have appealing statistical links, they are applied in VL and this paper, respectively, at a level of generality that transcends statistical contexts.

3. Examples

In this Section we list briefly versions of the ISRA for a variety of particular cases, drawing heavily from the examples considered by VL. In particular, we illustrate the performance of the ISRA on their 'cart' experiment, and compare the resulting restoration with that obtained with the EM algorithm.

Example 1. Inversion of simple linear equations

This example is extremely trivial, but illustrates elementary features of the behaviour of the ISRA and EM algorithms in solving under-determined linear systems.

Case 1. One equation in two unknowns

Suppose we wish to find f_1 and f_2 to solve

$$g = h_1 f_1 + h_2 f_2.$$

The ISRA is easily shown to be

$$f_i^{(n)} = f_i^{(n-1)} g / \{h_1 f_1^{(n-1)} + h_2 f_2^{(n-1)}\}, \quad i = 1, 2.$$

Thus, for any $f^{(o)}, h_1 f_1^{(1)} + h_2 f_2^{(2)} = g$, so that the ISRA converges in one step. It is easy to show that the same is true of the EM algorithm.

Case 2. Two equations in three unknowns

In this case, neither algorithm converges at once. For illustrative purposes, consider the equations

$$\begin{aligned} g_1 &= f_1 + f_2, \\ g_2 &= f_2 + f_3, \end{aligned}$$

with $g_1 = g_2 = 2$. Clearly any f of the form $f^T = (f_1, 2 - f_1, f_1)$ solves these equations. The ISRA is

$$f_1^{(n)} = 2f_1^{(n-1)}/(f_1^{(n-1)} + f_2^{(n-1)}),$$
$$f_2^{(n)} = 4f_2^{(n-1)}/(f_1^{(n-1)} + 2f_2^{(n-1)} + f_3^{(n-1)}),$$
$$f_3^{(n)} = 2f_3^{(n-1)}/(f_2^{(n-1)} + f_3^{(n-1)}).$$

The first and third equations for the EM algorithm match those of the ISRA, whereas the second one is

$$f_2^{(n)} = f_2^{(n-1)}\{(f_1^{(n-1)} + f_2^{(n-1)})^{-1} + (f_2^{(n-1)} + f_3^{(n-1)})^{-1}\}.$$

Table 1 compares the algorithms in terms of the number of iterations required to obtain the limiting point correct to three decimal places in each component of $\hat{f}$. If $f^{(o)} = (a, a, a)$, for any a, each algorithm converges at once to $\hat{f} = (1, 1, 1)$. If $f^{(o)} = (a, b, a)$, for $b \neq a$, each algorithm also converges at once, to the same $\hat{f}$ for both algorithms. Otherwise, the algorithms converge to different, but similar, $\hat{f}$'s, in roughly the same number of iterations as each other.

Table 1. Iterations required for convergence of ISRA and EM algorithms from various starting points. Solutions required to be correct to 3 decimal places in all elements.

$f^{(o)}$	$(\hat{f}_1, \hat{f}_2)$		Iterations	
	ISRA	EM	ISRA	EM
(0.5, 1.0, 1.5)	(0.951, 1.049)	(0.906, 1.094)	12	13
(0.5, 1.5, 1.0)	(0.646, 1.354)	(0.636, 1.364)	18	18
(1.5, 0.5, 1.0)	(1.425, 0.575)	(1.416, 0.584)	6	7
(0.1, 1.0, 9.9)	(0.648, 1.352)	(0.664, 1.336)	10	20
(1.0, 0.1, 9.9)	(1.963, 0.037)	(1.899, 0.101)	3	3
(1.0, 9.9, 0.1)	(0.063, 1.937)	(0.061, 1.939)	190	182

Example 2. Portfolio Optimization (VL, Section 3.1)

VL introduce a different notation here but we retain that of our earlier Sections. As a result, the algorithm is precisely that of (1.1), with the following interpretations for f, g and H. For each i, f_i is the proportion of total assets to be invested in stock i, the jth column of H contains one of the N possible sets of returns from the M stocks, and g_j denotes the probability that the jth set of returns will materialize, $j = 1, \ldots, N$.

This example is essentially the same as the example of grouped data considered in Section 3.2 of VL.

Example 3. Emission Tomography (VL, Section 3.3, Shepp and Vardi (1982))

Again (1.1) is the relevant ISRA, if we use f_i to denote the emission intensity at pixel i, g_j as the number of recorded events at detector j, and h_{ij} as the probability that a Poisson emission in pixel i is picked up in detector j. This manifestation is similar to the case of estimating the mixing weights of a mixture of M known multinomials, each defined on the same sample space of N categories. In that case, f is the set of mixing weights, g is the proportion of observations that fall into category j, and the ith row of H contains the ith 'pure' multinomial distribution.

Titterington and Rossi (1985) noted the relationship between these two problems in the context of the EM algorithm, building on the earlier work of Di Gesu and Maccarone (1984).

Example 4. Mixtures (VL, Section 3.4)

The special case of finite mixtures of multinomials is dealt with in Example 3. For more general finite mixtures, VL distinguish between two cases related to the "estimation" of the mixing weights f_i corresponding to the following version of (2.3):

$$G(\cdot) = \sum_i f_i H(i, \cdot). \tag{3.1}$$

Suppose $h_i(\cdot)$ denotes the density associated with $H(i, \cdot)$. If the problem is the statistical estimation of the $\{f_i\}$ from a random sample $Y_1, \ldots, Y_N$ from the mixture, and if G_N denotes the corresponding empirical distribution, then the ISRA is

$$f_i^{(n)} = f_i^{(n-1)} \Big\{ \int_{D_g} h_i(y) dG_N(y) \Big\} \Big/ \Big[\int_{D_g} h_i(y) \Big\{ \sum_s f_s^{(n-1)} h_s(y) \Big\} dy \Big]. \tag{3.2}$$

If, on the other hand, one is simply inverting (3.1), then the ISRA is given by (3.2) but with G_N replaced by G.

Example 5. Convolutions and Motion Blurring (VL, Sections 4 and 5).

In this example, $h(x, y) \equiv h(y - x)$. We follow VL in concentrating, for simplicity, on the case of one-dimensional images, so that $f(\cdot)$ and $g(\cdot)$ denote, respectively, the unblurred and blurred images, and $\Gamma = \{\gamma(t), 0 \leq t \leq T\}$ describes the blur in terms of the path followed by an origin of the coordinate system during the exposure interval $[0, T]$ of the photograph that produced g. Thus, instead of (2.3) we have, for all y,

$$g(y) = \int_0^T f\{y - \gamma(t)\} dt = \int f(x) \frac{1_\Gamma(y - x)}{|\gamma'\{\gamma^{-1}(y - x)\}|} dx,$$

where the limits of integration are defined by the indicator function. In terms of (2.3),

$$h(x,y) \equiv 1_\Gamma(y-x)/|\gamma'\{\gamma^{-1}(y-x)\}|.$$

In the continuous case (c.f. (4.8) of VL) the general form of the ISRA is

$$f^{(n)}(x) = f^{(n-1)}(x)\Big\{\int_{x+\Gamma} |\gamma'\{\gamma^{-1}(y-x)\}|^{-1} g(y)dy\Big\}\Big/ \Big[\int_{x+\Gamma} |\gamma'\Big\{\gamma^{-1}(y-x)\Big\}|^{-1}\Big\{\int_{y-\Gamma} f^{(n-1)}(s)|\gamma'\{\gamma^{-1}(y-x)\}|^{-1}ds\Big\}dy\Big].$$

Special versions for the cases of constant-speed linear motion and constant acceleration from rest along a straight line (both dealt with by VL) are easily written down (see Archer and Titterington (1993) for details).

In the discrete case, the general form of the algorithm is given by (1.1). In the version for constant-speed linear motion, $h_{ij} = a1_{\{0,\ldots,b\}}(j-i)$, for some $a > 0$, and (1.1) becomes (c.f. (5.6) of VL)

$$f_i^{(n)} = f_i^{(n-1)} a^{-1}\Big(\sum_{j=i}^{i+b} g_j\Big)\Big/\Big\{\sum_{j=1}^{i+b}\sum_{s=L_1}^{L_2} f_s^{(n-1)}\Big\}, \tag{3.3}$$

where $L_1 = \max\{1, j-b\}$ and $L_2 = \min\{M, j\}$.

For details of the version for constant acceleration from rest along a straight line, see Archer and Titterington (1993).

We applied the algorithm to the ‘cart’ example that constitutes *Experiment* 2 of VL. We followed their procedure as closely as possible and ran both the EM and ISRA algorithms for 106 iterations.

Two details of the image were considered, each of which was a 250×250 pixel scene ($M = 62500$). It was assumed, as in VL, that motion blur of 106 pixels had been imposed and, in analyzing the two subimages, data were used from adjoining strips of widths 106 pixels. The ISRA was therefore based on (3.3).

Figure 1(a) shows the first sub-image, which comprises the area of the whole image near a blurred wheel. Figs. 1(b) and 1(d) show, respectively, the results of applying the EM-algorithm and the ISRA for 106 iterations, and Fig. 1(c) shows the difference-image between EM and ISRA. The only substantial differences occur near the bottom left edge of the wheel. In both Fig. 1(b) and Fig. 1(d) there is evidence of the vertical artefacts, mentioned in VL, that are, interestingly, less evident after 40 iterations, as shown in Fig. 2. Fig. 2 is the 40-iterations equivalent of Fig. 1; at that stage, the EM and ISRA results are very similar.

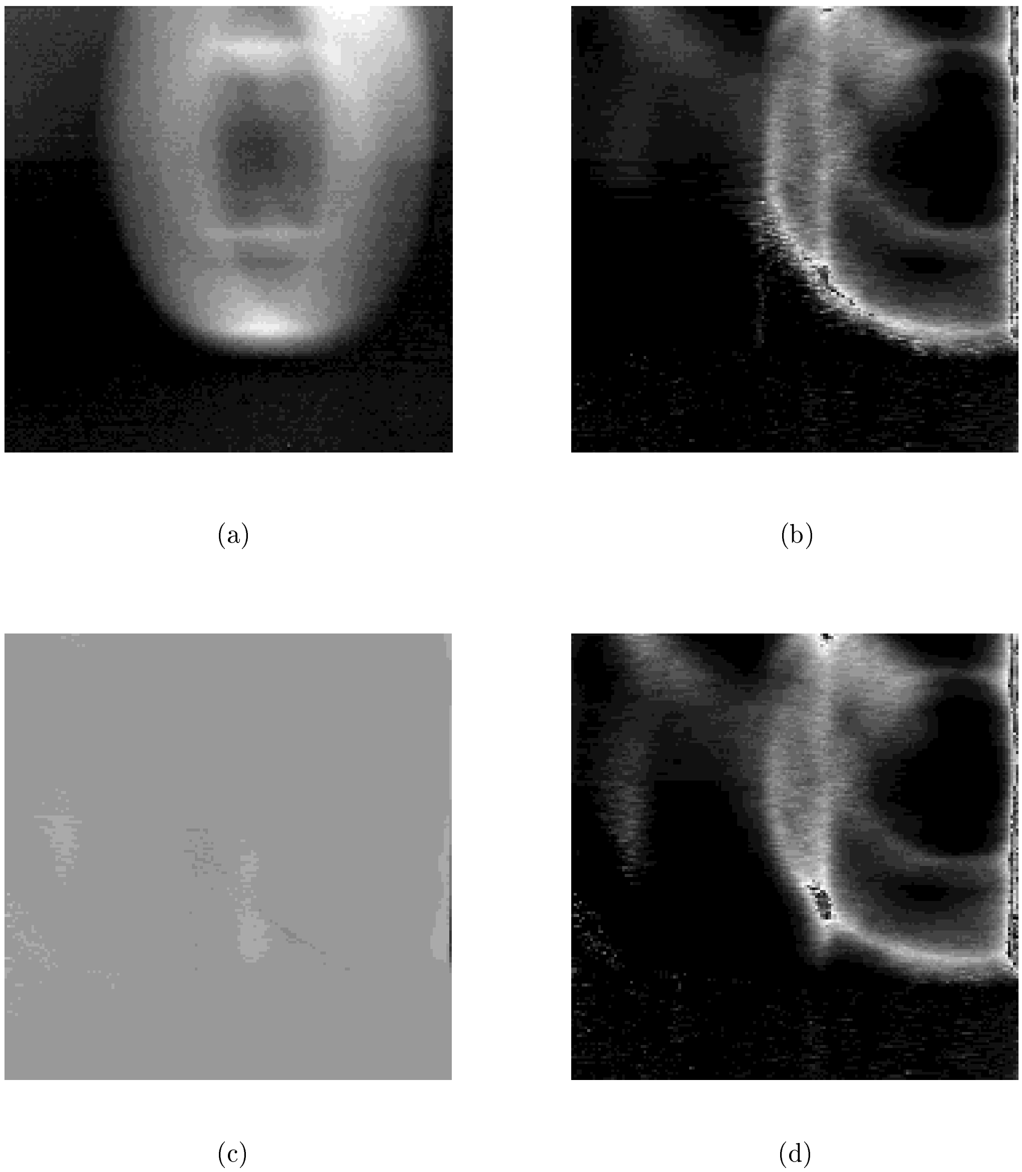


Figure 1. Wheel sub-image after 106 iterations: (a) data; (b) EM; (c) EM-ISRA difference image; (d) ISRA.

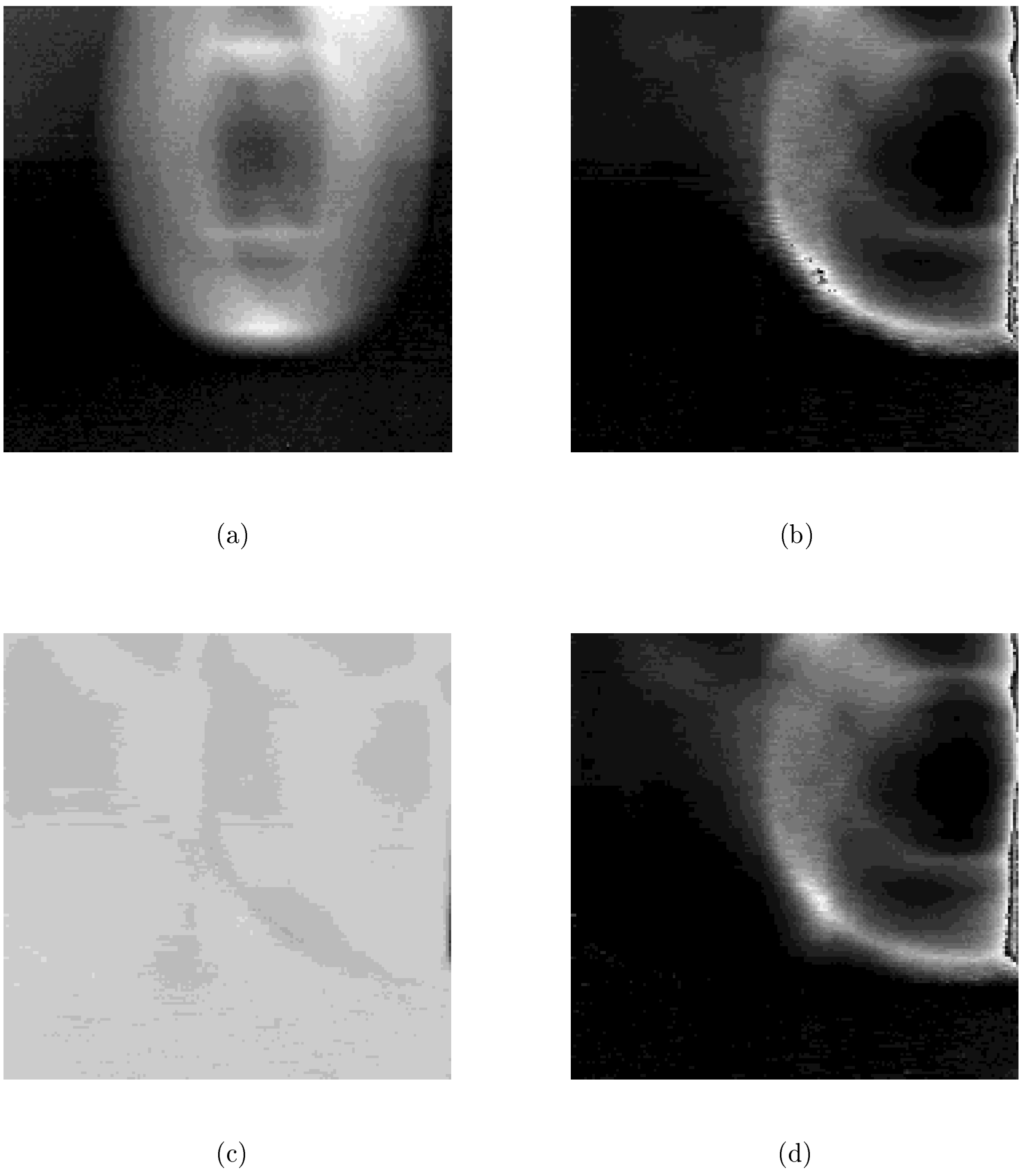


Figure 2. Wheel sub-image after 40 iterations: (a) data; (b) EM; (c) EM-ISRA difference image; (d) ISRA.

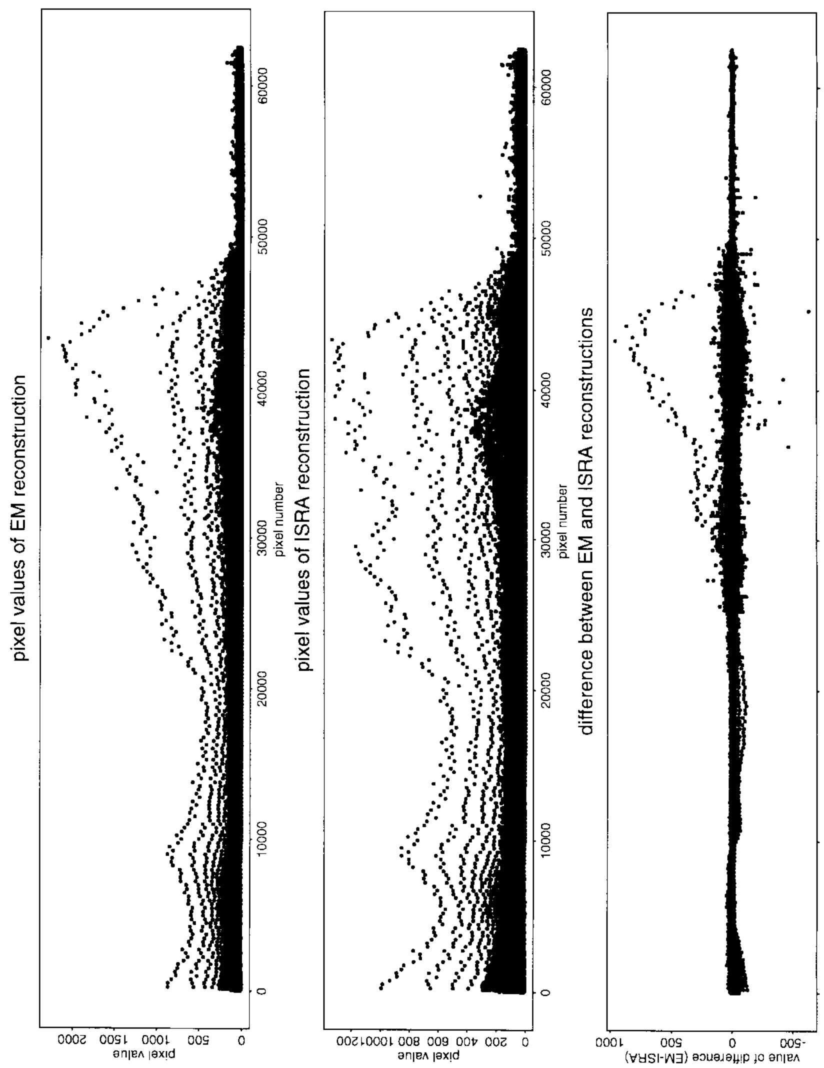
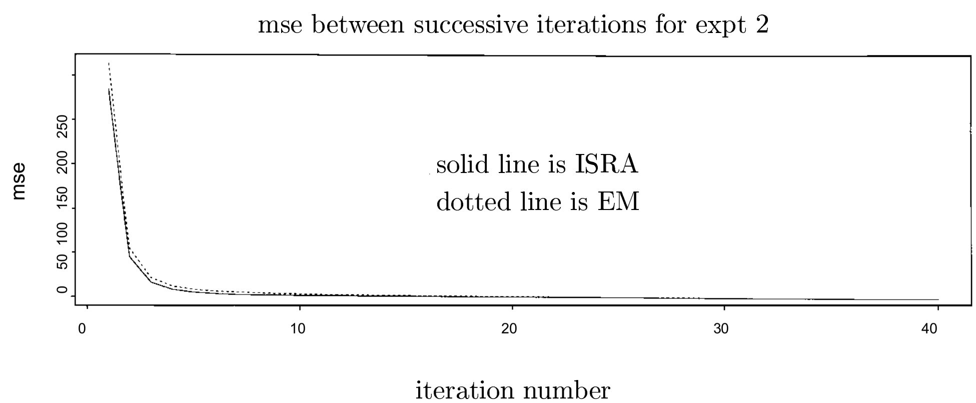


Figure 3. Pixel intensities for wheel sub-image after 106 iterations: (a)EM; (b) ISRA; (c) difference image.



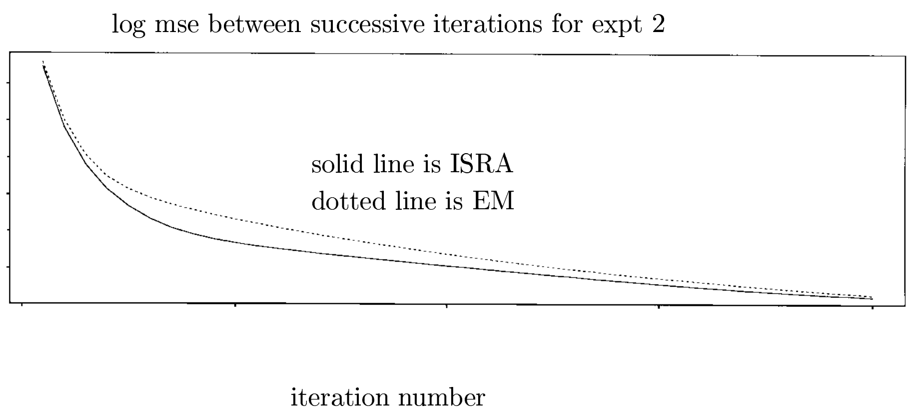
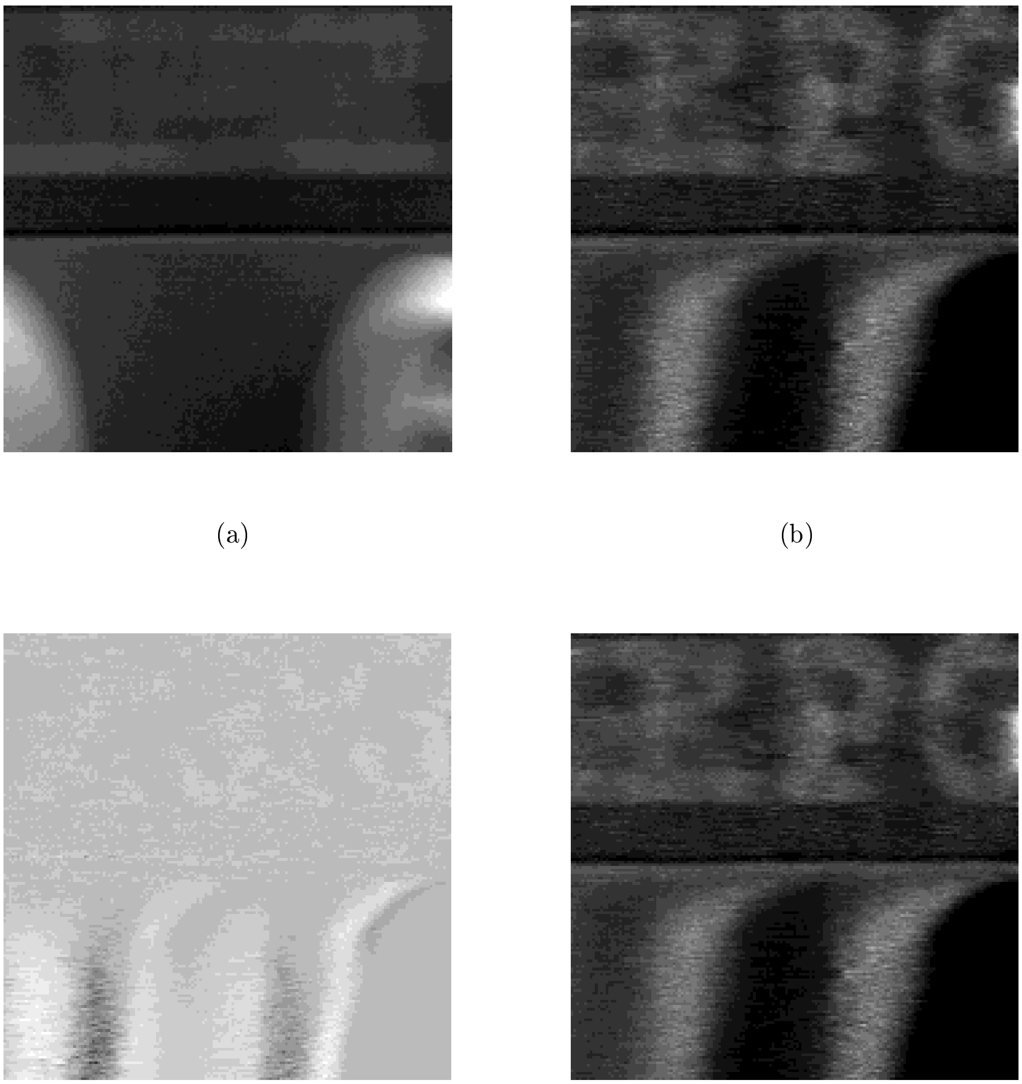


Figure 4. Mean squared differences between successive iterates for wheel sub-image using EM(——) and ISRA (.......): (a) actual values; (b) log-scale.

Fig. 3 plots the pixel intensities corresponding to Fig. 1(b) (Fig. 3(a)), Fig. 1(d) (Fig. 3(b)) and Fig. 1(c) (Fig. 3(c)), where the pixels are numbered by raster scan beginning at the top left-hand corner. Fig. 4 gives some idea of the convergence behaviours of the two algorithms, in terms of the per-pixel mean-squared difference between successive iterates:

$$M^{-1}\sum_{i=1}^{M}(f_i^{(n)} - f_i^{(n-1)})^2.$$

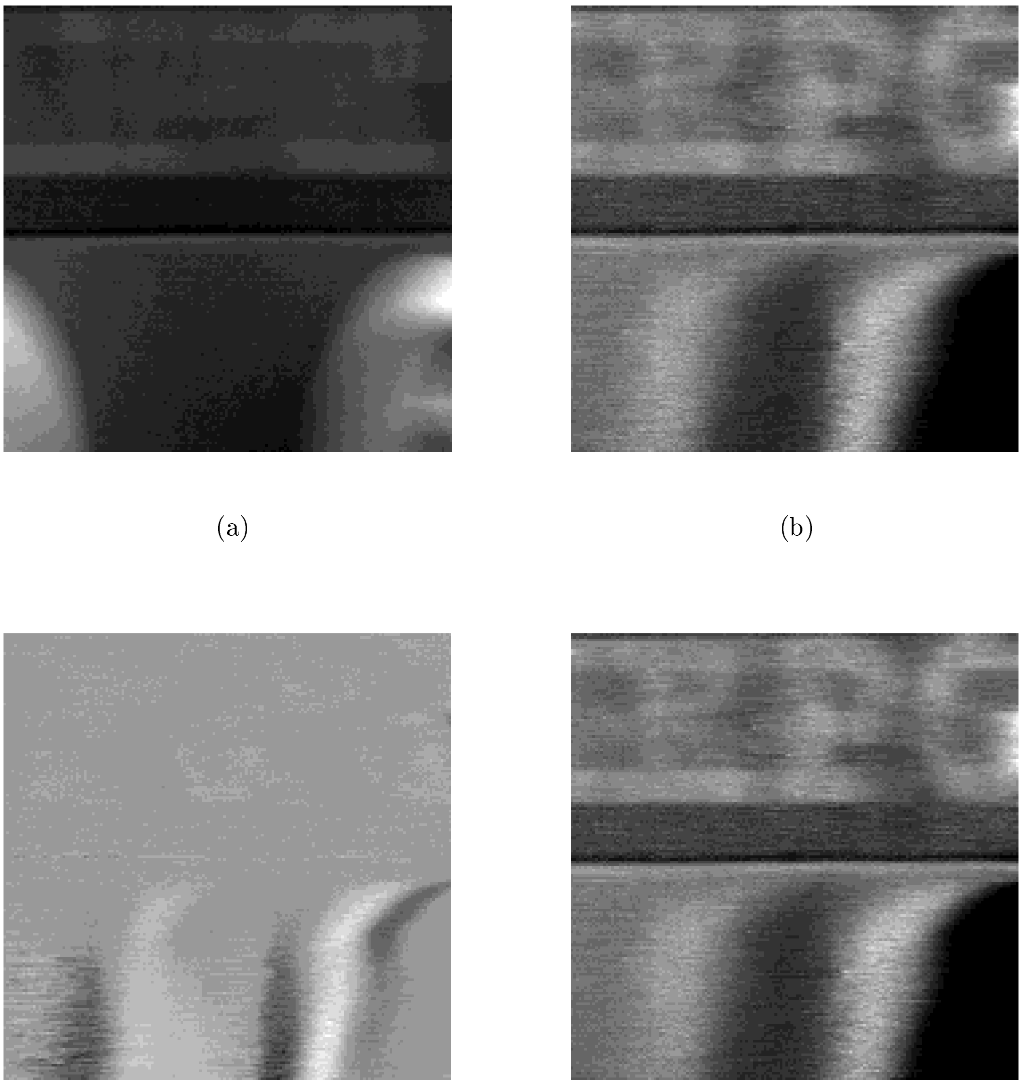
Both Fig. 4(a) and, on the log-scale, Fig. 4(b), indicate that the behaviour is very similar, with EM consistently taking slightly larger steps.



(a) (b) (c) (d)

Figure 5. Letter sub-image after 106 iterations: (a) data; (b) EM; (c) EM-ISRA difference image; (d) ISRA.

The other part of the image examined includes the letters ‘RPO’ from the word ‘AIRPORT’. Fig. 5 is the 106-iterate version (c.f. Fig. 1) and Fig. 6 corresponds to 40 iterations (c.f. Fig. 2). Again, the two algorithms produced,



(a) (b)

(c) (d)

Figure 6. Letter sub-image after 40 iterations: (a) data; (b) EM; (c) EM-ISRA difference image; (d) ISRA.

in general, very similar results, and the 40-iterate restorations seem to be at least as appealing as those after 106 iterations. It seems likely that, as emphasized later in Section 5.2, the inverse problem is somewhat ill-posed, and that stopping at, say, 40 iterations imposes beneficial regularization.

4. Convergence of the ISRA

As remarked by VL (end of Section 1.3) in the context of their algorithm, if the LININPOS has no nonnegative solution (in particular, if $M < N$ in the discrete version), then convergence of the algorithm should obtain to a closest, in some sense, approximation to such a solution.

Convergence aspects of the discrete ISRA are discussed by Titterington (1987) and De Pierro (1987, 1990), and for Chahine's algorithm by Chu (1985). However, the most complete discussion appears to be that of Eggermont (1990) and we summarize this below.

Define $D_r = \mathrm{diag}\{f_i^{(r)}/(HH^T f^{(r)})_i, i = 1, \ldots, M\}$. Then, by Lemma 6.1 of Eggermont (1990),

$$S(f^{(n-1)}) - S(f^{(n)}) \geq (f^{(n-1)} - f^{(n)})^T D_{n-1}^{-1} (f^{(n-1)} - f^{(n)}), \tag{4.1}$$

where $S(f)$ is defined in (2.6). Thus, $\{S(f^{(n)})\}$ is decreasing and since, for every n, $\{f \geq 0 : S(f) \leq S(f^{(o)})\}$ is a compact set, $\{f^{(n)}\}$ is bounded and every subsequence itself contains a convergent subsequence.

Now let $\hat{f}$ be any point of accumulation of $\{f^{(n)}\}$, let $\Lambda = \mathrm{diag}\{(Hg)_i, i = 1, \ldots, M\}$, let

$$\Delta_{\mathrm{KL}}(x; y) = \sum_{i=1}^{M} \{x_i \log(x_i/y_i) + y_i - x_i\},$$

the Kullback-Leibler directed divergence for $x > 0, y > 0$, and let

$$e(\hat{f}; f) = \Delta_{\mathrm{KL}}(\Lambda\hat{f}; \Lambda f) + S(f) - S(\hat{f}). \tag{4.2}$$

Then Lemma 6.2 of Eggermont (1990) shows that

$$e(\hat{f}; f^{(n-1)}) \geq e(\hat{f}; f^{(n)}).$$

The convergence properties of $\{f^{(n)}\}$ imply that $\{e(\hat{f}; f^{(n)})\}$ and, hence, $\{\Delta_{\mathrm{KL}}(\Lambda\hat{f}; \Lambda f^{(n)})\}$ converge to zero. The last result implies that $\{f^{(n)}\}$ converges to $\hat{f}$, which can be shown to be a minimizer of $S(f)$, subject to $f \geq 0$. The consummation of this analysis is therefore the following theorem.

Theorem 1. *The* ISRA *defined by* (1.1) *and initialized by* $f^{(o)} > 0$, *generates a sequence* $\{f^{(n)}\}$ *that converges to an* $\hat{f}$ *that minimizes* $S(f)$ *subject to* $f \geq 0$.

Extension beyond the discrete case proceeds along similar lines to those in Section 3.4 of VL; the main steps are sketched below.

4.1. The case of finite D_f

Suppose $D_f = \{1, \dots, M\}$, that

$$g(y) = \sum_{i=1}^{M} h_i(y) f_i, \quad y \in D_g,$$

and that $\{f^{(n)}\}$ are generated according to

$$f_i^{(n)} = f_i^{(n-1)} \Big\{ \int_{D_g} h_i(y) g(y) dy \Big\} \Big/ \Big[\int_{D_g} h_i(y) \Big\{ \sum_{s=1}^{M} f_s^{(r)} h_s(y) \Big\} dy \Big], \ i = 1, \dots, M, \tag{4.3}$$

starting from $f^{(o)} > 0$. Then (4.1) and (4.2) hold with

$$S(f) = \int_{D_g} \Big\{ g(y) - \sum_i h_i(y) f_i \Big\}^2 dy, \tag{4.4}$$

$$D_r = \operatorname{diag}\Big\{ f_i^{(r)} \Big/ \Big[\int_{D_g} h_i(y) \Big\{ \sum_{s=1}^{M} f_s^{(r)} h_s(y) \Big\} dy \Big], \quad i = 1, \dots, M \Big\}$$

and $\Lambda = \operatorname{diag}\{\int_{D_g} h_i(y) g(y) dy, i = 1, \dots, M\}$. The argument of Theorem 1 then confirms that $\{f^{(n)}\}$ converges to an $\hat{f}$ that minimizes $S(f)$, as defined by (4.4), subject to $f \geq 0$.

4.2. The case of continuous D_f

As in the case of Section 3.4.2 of VL, we attack this by way of a discretization approach. Suppose $g(\cdot)$ and $h(x, \cdot)$ are nonnegative, integrable functions on D_g and suppose

$$g(y) = \int_{D_f} h(x, y) f(x) dx, \quad x \in D_f, \tag{4.5}$$

has a non-negative solution that is piecewise constant over the measurable partition $\{B_1, \dots, B_{M'}\}$ of $D_f (\mu(B_i) > 0$, all $i)$. Then, for any refinement $\{A_1, \dots, A_M\}$ of $\{B_1, \dots, B_{M'}\} (\mu(A_i) > 0$, all $i)$, the ISRA

$$\lambda_i^{(n)} = \lambda_i^{(n-1)} \Big\{ \int_{D_g} H_2(A_i, y) g(y) dy \Big\} \Big/ \Big[\int_{D_g} H_2(A_i, y) \Big\{ \sum_s \lambda_s^{(n-1)} H_2(A_s, y) \Big\} dy \Big], \tag{4.6}$$

$i = 1, \dots, M$, $n = 1, 2, \dots$, initialized by $\lambda^{(o)} > 0$, converges to a limiting value $\lambda^* \ (\geq 0)$, and

$$f_{M^*}(x) = \sum_{i=1}^{M} \lambda_i^* 1_{A_i}(x), \qquad x \in D_f, \tag{4.7}$$

is a solution of (4.5). In (4.6), $H_2(A_i, y) = \int_{A_i} h(x, y)dx$, all i, and in (4.7), $1_{A_i}(\cdot)$ is the indicator function for A_i, all i.

This result follows from the results already obtained in this, our Section 4, and the final steps to the continuous case are parallel to those in Section 3.5 of VL.

5. Discussion

5.1. ISRA versus EM

The illustration in Section 3 provides empirical information about the comparison between the ISRA and the EM algorithm. In general, both algorithms are prone to slow convergence. In the context of emission tomography (ET), Daube-Witherspoon and Muehllehner (1986) provide further empirical evidence of this, but point out that the number of operations per iteration is significantly lower in the ISRA. Both algorithms can be accelerated, either by applying Aitken's Δ^2 procedure, or by adding a linear search embellishment. Lewitt and Muehllehner (1986) implement this in the case of the EM algorithm for ET, and De Pierro (1987) points out for the ISRA that both the optimal step-length in the proposed direction is easily computed and the link with Chahine's (1970) algorithm makes available further improvement mechanisms developed in other branches of the inverse-problems literature.

An alternative general approach to the inversion of linear equations is through the Fourier domain. In the situation when there are missing data, Ollinger and Karp (1988) compare the ISRA with two such methods, finding that the ISRA is slow in comparison but admitting that it could be accelerated.

The (very limited) evidence from Table 1 is that the convergence rates of the EM and ISRA algorithms may be similar. The local (near the limit point) convergence behaviour in the case of f with finite domain, D_f, is dependent on Ostrowski's Theorem; see, for instance, Ortega and Rheinboldt (1970, p.300). Consider an iterative algorithm of the form

$$f_i^{(n)} = \phi_i(f^{(n-1)}), \quad i = 1, \ldots, M, \quad n = 1, 2, \ldots,$$

and suppose that $\hat{f}$ is the limit of $\{f^{(n)}\}$. Define the matrix $U(f) = \{U_{is}(f)\}$ by $U_{is}(f) = \partial\phi_i(f)/\partial f_s$, for $i, s = 1, \ldots, M$. Then the rate of local convergence to $\hat{f}$ is dictated by the spectral radius of $U(\hat{f})$.

Consider now the versions of $U(f)$ corresponding to the discrete ISRA given by (1.1) and the discrete EM algorithm defined by

$$f_i^{(n)} = f_i^{(n-1)}\Big(\sum_{j=1}^{N} h_{ij}\Big)^{-1} \sum_{j=1}^{N}\Big\{h_{ij}\Big/\Big(\sum_{s=1}^{M} f_s^{(n-1)}h_{sj}\Big)\Big\}g_j,\ i = 1, \ldots, M. \tag{5.1}$$

This is the version of (1.2) corresponding to the more general case in which $\sum_j h_{ij} \neq 1$ for all i. It is straightforward (Archer and Titterington (1993)) to show that, for the ISRA,

$$U_{is}(\hat{f}) = \delta_{is} - \hat{f}_i \Big(\sum_{j=1}^{N} h_{ij} h_{sj}\Big) \Big/ \Big(\sum_{j=1}^{N} h_{ij} g_j\Big), \quad i, s = 1, \ldots, M, \tag{5.2}$$

and, for the EM algorithm,

$$U_{is}(\hat{f}) = \delta_{is} - \hat{f}_i \Big(\sum_{j=1}^{N} h_{ij}\Big)^{-1} \sum_{j=1}^{N} \Big(h_{ij} h_{sj} / g_j\Big), \quad i, s = 1, \ldots, M, \tag{5.3}$$

in both of which δ_{is} is the Kronecker delta. The methods of Titterington (1987) show that, in both cases, the eigenvalues of $U(\hat{f})$ are non-negative and strictly less than unity. Comparison of the maximum eigenvalues in particular cases would complete the comparison of local convergence properties. In the case of the illustration in Example 1, Case 2, in which $N = 2$ and $g_1 = g_2$, (5.2) and (5.3) are identical.

5.2. Addition of penalty functions

Many practical inverse problems suffer from ill-posedness, with the result that direct inversion can lead to an unsatisfactory $\hat{f}$, particularly if there is any chance of uncertainty or noise. A common device for counteracting this is to add a roughness penalty function to a loglikelihood or sum-of-squares function before optimizing. The resulting $\hat{f}$ is typically improved in terms of mean-squared error properties and can often be interpreted in Bayesian terms as a maximum a posteriori (MAP) estimate of f.

Consider, first, the discrete case. Instead of $S(f)$ in (2.6) one would consider

$$S_{\lambda,C}(f) = \|g - H^T f\|_2^2 + \lambda f^T C f, \tag{5.4}$$

in which $\lambda > 0$ and C is nonnegative definite. Since (5.4) is interpretable as equivalent to the logarithm of a (Normal) posterior density for f, given g, an EM algorithm for seeking the posterior mode (the MAP estimate of f) can be constructed along the lines of Section 4.5 of Dempster et al. (1977).

So far as the ISRA is concerned, any minimizer of $S_{\lambda,C}(f)$ satisfies

$$Kf = v,$$

where $K = (HH^T + \lambda C)$ and $v = Hg$, stimulating the algorithm

$$f_i^{(n)} = f_i^{(n-1)} v_i \Big/ \Big(\sum_s k_{is} f_s^{(n-1)}\Big), \qquad i = 1, \ldots, M.$$

In this case, it is typical that some of the elements of C are negative, so that it is not automatic that $f^{(n)} \geq 0$ for all n.

In the case of a continuous function, the usual kind of penalty function is based on derivatives of f, a common one being

$$\lambda \int_{D_f} \{f''(x)\}^2 dx, \tag{5.5}$$

where $f(\cdot)$ is the density associated with the measure $F(\cdot)$. However, it is common to restrict the choice of f to some space spanned by a certain class of basis functions, and (5.5) reduces to a quadratic form in a transformed, finite vector of parameters (see Silverman (1985), for example). The corresponding $S(f)$ is similarly transformed and the problem reverts to one of the form (5.4). Byrne (1993) studies penalized versions of the EM algorithm.

5.3. Wider Classes of Algorithms

In this final section we refer again to Eggermont (1990), who considers wider classes of algorithms which include the EM algorithm and the ISRA as special cases. Consider the (discrete-case) problem of minimizing $S(f)$, subject to $f \geq 0$, where $S(\cdot)$ is a convex, continuously differentiable function on $\mathbb{R}^M$, with compact level sets and locally Lipschitz continuous gradient. Eggermont (1990) considers three classes of multiplicative, iterative algorithms, of the following forms:

$$f_i^{(n)} = f_i^{(n-1)}[1 - w_n\{\nabla S(f^{(n-1)})\}_i],\ i = 1, \ldots, M, \tag{5.6}$$

$$f_i^{(n)}[1 + w_n\{\nabla S(f^{(n)})\}_i] = f_i^{(n-1)}, \quad i = 1, \ldots, M, \tag{5.7}$$

and

$$f_i^{(n)} = f_i^{(n-1)}/[1 + w_n\{\nabla S(f^{(n-1)})\}_i],\ i = 1, \ldots, M. \tag{5.8}$$

In (5.6)-(5.8), w_n is a step-length parameter. Algorithm (5.7) is called an implicit algorithm and (5.8) explicit. For appropriate choices of S, $w_n \equiv 1$ in (5.6) gives the EM algorithm in emission tomography and (5.8) gives the ISRA. In discussing (5.8), Eggermont (1990) develops the convergence properties for the ISRA as described above in Section 4, and establishes the convergence properties of the implicit algorithm (5.7) by a similar but slightly simpler argument.

It would be of interest to investigate other versions of these algorithms, for different choices of S, for various choices of the step-lengths $\{w_n\}$, and in the context of the versions appropriate for solving integral equations.

Acknowledgements

The authors are grateful to Professor P. J. Green for helpful discussions about gaining access to the 'cart' data. One of us (G.E.B.A.) was supported by a Research Studentship from the UK Science and Engineering Research Council.

References

Archer, G. E. B. and Titterington, D. M. (1993). On the iterative Image Space Reconstruction Algorithm (ISRA) for solving linear inverse problems. Technical Report 93-2, Department of Statistics, University of Glasgow, Scotland.

Byrne, C. L. (1993). Iterative image reconstruction algorithms based on cross-entropy minimization. IEEE Trans. Image Process. **2**, 96-103.

Chahine, M. T. (1970). Inverse problems in radioactive transfer: Determination of atmospheric parameters. J. Atmos. Sci. **27**, 960-967.

Chu, W. P. (1985). Convergence of Chahine's nonlinear relaxation inversion method used for limb viewing remote sensing. Appl. Optics **24**, 445-447.

Daube-Witherspoon, M. E. and Muehllehner, G. (1986). An iterative image space reconstruction algorithm suitable for volume ECT. IEEE Trans. Med. Imgng. **5**, 61-66.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. Ser.B **39**, 1-38.

De Pierro, A. R. (1987). On the convergence of the iterative Image Space Reconstruction Algorithm for volume ECT. IEEE Trans. Med. Imgng. **6**, 174-175.

De Pierro, A. R. (1990). Nonlinear relaxation methods for solving symmetric linear complementarity problems. J. Optim. Theory Appl. **64**, 87-99.

Di Gesu, V. and Maccarone, M. C. (1984). The Bayesian direct deconvolution method: Properties and applications. Signal Process. **6**, 201-211.

Eggermont, P. P. B. (1990). Multiplicative iterative algorithms for convex programming. Lin. Alg. Appl. **130**, 25-42.

Lewitt, R. M. and Muehllehner, G. (1986). Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimators. IEEE Trans. Med. Imgng. **5**, 16-22.

Ollinger, J. M. and Karp, J. S. (1988). An evaluation of three algorithms for reconstructing images from data with missing projections. IEEE Trans. Nucl. Sci. **35**, 629-634.

Ortega, J. M. and Rheinboldt, W. C. (1970). Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. IEEE Trans. Med. Imgng. **1**, 113-122.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). J. Roy. Statist. Soc. Ser.B **47**, 1-52.

Titterington, D. M. (1987). On the iterative Image Space Reconstruction Algorithm for ECT. IEEE Trans. Med. Imgng. **6**, 52-56.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). Statistical Analysis of Finite Mixture Distributions. John Wiley, Chichester.

Titterington, D. M. and Rossi, C. (1985). Another look at a Bayesian direct deconvolution method. Signal Process. **9**, 101-106.

Vardi, Y. and Lee, D. (1993). From image deblurring to optimal investments: Maximum likelihood solutions for positive linear inverse problems (with discussion). J. Roy. Statist. Soc. Ser.B **55**, 569-612.

Department of Statistics, University of Glasgow, Glasgow, G12 8QQ, Scotland.

(Received June 1993; accepted July 1994)