Journal of X-Ray Science and Technology 10 (2002) 77–86
IOS Press

Development of iterative algorithms for image reconstruction

Ming Jiang[a,*] and Ge Wang[b]
[a]*School of Mathematics, Peking University, Beijing 100871, China*
E-mail: jiangm@math.pku.edu.cn
URL: http://ct.radiology.uiowa.edu/˜jiangm
[b]*Department of Radiology, University of Iowa, Iowa City, IA 52242, USA*
E-mail: ge-wang@ieee.org
URL: http://dolphin.radiology.uiowa.edu/ge

Abstract. The iterative approach is important for computed tomography (CT) and attracting more and more attention with the rapid evolution of computer technology. In this paper, classic results and recent advances on iterative algorithms for image reconstruction are reviewed, with an emphasis on the ART-like and EM-like algorithms in both of their simultaneous and ordered-subset formats. The following issues are discussed: what the computational structures are, under what conditions the algorithms converge, what the final limits are, what the relaxation strategies we have, how the final limits depend on the initial guesses, and so on. In addition, heuristic arguments are given for the SART and EM algorithms for understanding the algorithms. Finally, future research directions are discussed, along with guidelines for practical applications of iterative algorithms.

1. Introduction

Image reconstruction plays a major role in many applications. Typically, an imaging device can be modeled by the following equation:

$$Ax = b \tag{1}$$

where the observed data $b = (b^1, \ldots, b^M) \in R^M$, original image $x = (x_1, \ldots, x_M) \in R^N$, and $A = (A_{ij})$ is a non-zero $M \times N$ matrix. The problem is to reconstruct the image x from the data b. A direct solution is not feasible with conventional direct methods because of the ill-posedness of the problem, the noisy corrupt data b and huge data dimension in practice.

The iterative approach has been important because of their superior performance in the above context. With the rapid development of computer technology, iterative algorithms receive increasingly more attention. Iterative algorithms may be categorized according to their criteria and the ways for updating an intermediate image with observed data. Generally, the criterion is to minimize either the least squares function or the I-divergence, which is respectively equivalent to maximize the likelihood in the Gaussian

*Corresponding author: Ming Jiang, PhD, School of Mathematics, Peking University, 5 Summer Palace Street, Beijing 100871, China. Tel.: +86 10 6275 6797; Fax: +86 10 6275 1801; E-mail: jiangm@ct.radiology.uiowa.edu or jiangm@math.pku.edu.cn; URL: http://ct.radiology.uiowa.edu/˜jiangm/.

or Poisson environments. The algorithms for least squares minimization are typically linear, while the algorithms for I-divergence minimization are nonlinear. As for the updating methods, projection data may be utilized simultaneously or block-wisely.

The algebraic reconstruction technique (ART) [13] and expectation maximization (EM) algorithm [20] are the primary algorithms widely used in the community, due to their simplicity, efficiency and performance. As will be clear later, the criteria that led to these two algorithms, either explicitly or implicitly stated, are the least squares function and the I-divergence, respectively. In the last decades, several improved algorithms were designed based on the ART and EM algorithms. One remarkable advance is the order-subset (OS) or block iterative (BI) versions of their simultaneous schemes. *"Recently, the OSEM approach has been adopted by the manufacturers of a number nuclear medicine imaging systems [18]"*. It has been found that *"the OS-EM algorithm provides an order-of-magnitude acceleration over EM, with restoration quality maintained [14]"*.

This paper is intended as a practical yet in-depth overview of both classic results and recent advances on iterative algorithms originated from the ART and EM algorithms. Both algorithmic and analytic aspects of the algorithms are appropriately addressed for the readers at large. The organization of the paper is as follows. After this introduction, Section 2 describes the ART and ART-like iterative algorithms, and unifies these specific algorithms within a general Landweber scheme. Then, the later part of Section 2 discusses the convergent properties of those schemes. Section 3 presents the EM algorithm and its OS variants, along with their properties. Section 4 formulates the SART and EM algorithms in a heuristic manner. Section 5 discuss relevant issues, suggests some research areas and concludes the paper.

2. ART-like algorithms and a general Landweber scheme

2.1. ART-like algorithms

ART: The ART is the first iterative algorithm used in CT [13]:

$$x_j^{(n+1)} = x_j^{(n)} + \lambda_n \frac{A_{ij}}{\|A^i\|^2} \left(b^i - A^i x^{(n)} \right) \tag{2}$$

where $i = n \bmod(M) + 1$, $\|A^i\|^2 = \sum_{j=1}^{N} A_{ij}^2$ is the Euclidean norm of the i-th row of A. This method was originally discovered by Kaczmarz in [17].

SART: In 1984, the SART was proposed as a major refinement of the ART [1]. Let

$$A_{i,+} = \sum_{j=1}^{N} |A_{ij}|, \quad \text{for } i = 1, \ldots, M, \tag{3}$$

$$A_{+,j} = \sum_{i=1}^{N} |A_{ij}|, \quad \text{for } j = 1, \ldots, N. \tag{4}$$

The SART is:

$$x_j^{(n+1)} = x_j^{(n)} + \lambda_n \frac{1}{A_{+,j}} \sum_{i=1}^{M} \frac{A_{ij}}{A_{i,+}} \left(b^i - A^i x^{(n)} \right). \tag{5}$$

Cimmino's Algorithm: The Cimmino's simultaneous projection method [10] is:

$$x_j^{(n+1)} = x_j^{(n)} + \lambda_n \frac{1}{M} \sum_{i=1}^{M} \frac{A_{ij}}{\|A^i\|^2} \left(b^i - A^i x^{(n)}\right). \tag{6}$$

DWE and CAV: In [7,8], an iterative scheme utilizing the sparsity of the matrix A was proposed. Consider a set $\{G_i\}_{i=1}^{M}$ of real diagonal $N \times N$ matrices:

$$G_i = \text{diag}\,(g_{i1}, \ldots, g_{iN}), \quad \text{with } g_{ij} \geqslant 0 \text{ such that } \sum_{i=1}^{M} G_i = I.$$

$\{G_i\}_{i=1}^{M}$ is referred to as being sparsity pattern oriented (SPO) with respect to an $M \times N$ matrix A if, for $i = 1, \ldots, M$ and $j = 1, \ldots, N$, $g_{ij} = 0$, if and only if $A_{ij} = 0$. The diagonal weighting (DWE) is the following iterative scheme:

$$x_j^{(n+1)} = x_j^{(n)} + \lambda_n \cdot \sum_{\substack{1 \leqslant i \leqslant M \\ g_{ij} \neq 0}} A_{ij} \frac{b^i - A^i x^{(n)}}{\sum_{\substack{1 \leqslant l \leqslant N \\ g_{il} \neq 0}} \frac{A_{il}^2}{g_{il}}}. \tag{7}$$

Censor et al. described a method to construct the SPO matrices $\{G_i\}_{i=1}^{M}$ according to the sparsity of A in [8]. Let

$$g_{ij} = \begin{cases} \frac{1}{s_j}, & \text{if } A_{ij} \neq 0, \\ 0, & \text{if } A_{ij} = 0, \end{cases} \tag{8}$$

where s_j is the number of nonzero elements $A_{ij} \neq 0$ in the j-th column of A. The component-averaging algorithm (CAV) is derived from Eq. (7) as

$$x_j^{(n+1)} = x_j^{(n)} + \lambda_n \cdot \sum_{1 \leqslant i \leqslant M} A_{ij} \frac{b^i - A^i x^{(n)}}{\sum_{1 \leqslant l \leqslant N} s_l A_{il}^2}. \tag{9}$$

2.2. General Landweber scheme

In [16], the following general Landweber scheme was studied for both of its simultaneous and OS versions. Let V and W be two positive definite diagonal matrices of order N and M, respectively. Most of the iteration algorithms to solve Eq. (1) can be written in the following form:

$$x^{(n+1)} = x^{(n)} + \lambda_n V^{-1} A^* W \left(b - Ax^{(n)}\right) \tag{10}$$

where $\lambda_n > 0$ is the relaxation coefficient. We use * to denote the transpose of a vector or matrix. The iterative scheme Eq. (10) is called simultaneous, which uses all the components of the observed data b for updating the iteration.

Let L_W be the following weighted least square functional

$$L_W(x) = \frac{1}{2}\|b - Ax\|_W^2 \tag{11}$$

for $x \in R^N$, where $\|\cdot\|_W$ is the norm induced on R^M by the following inner product associated with W: $\langle x, y\rangle_W = \langle Wx, y\rangle = \sum_{i=1}^{M} W_i x_i y_i$. As we will explain more below, the sequence $\{x^{(n)}\}$ generate by Eq. (10) converges to a minimizer of L_W under certain conditions. All the minimizers of L_W satisfy the following normal equation $\nabla L_W(x) = \theta$:

$$A^* W A x = A^* W b. \tag{12}$$

If there is a solution of Eq. (1), it is called consistent. Otherwise, it is called inconsistent. If Eq. (1) is consistent, it can be shown that Eqs (1) and (12) have exactly the same solutions. Hence, the minimal norm solutions of Eqs (1) and (12) are the same in the consistent case.

In recent years, great interests are in the OS version of Eq. (10), which invokes only part of the observed data b each time. Assume that the index set $B = \{1, \ldots, M\}$ is partitioned into T nonempty subsets B_t such that

$$B = \{1, \ldots, M\} = \bigcup_{1 \leqslant t \leqslant T} B_t. \tag{13}$$

For simplicity, we assume that these subsets B_i are disjoint, although this requirement is not necessary in general, cf. [16] for more details. Assume

$$B_t = \{i_1^t, \ldots, i_{M(t)}^t\} \tag{14}$$

In the following, let $[n] = n \operatorname{mod}(T) + 1$ for $n \geqslant 0$. Let

$$A_t = \begin{pmatrix} A^{i_1^t} \\ \vdots \\ A^{i^t_{M(t)}} \end{pmatrix}_{M(t)\times N}, \quad W_t = \begin{pmatrix} W^{i_1^t} & & \\ & \ddots & \\ & & W^{i^t_{M(t)}} \end{pmatrix}_{M(t)\times N}, \quad b_t = \begin{pmatrix} b^{i_1^t} \\ \vdots \\ b^{i^t_{M(t)}} \end{pmatrix} \in R^{M(i)}.$$

where A^i, W^i and b^i are the i-th row of A, the i-th diagonal element of W, and the i-th component of b, respectively. The OS version of Eq. (10) can be written as

$$x^{(n+1)} = x^{(n)} + \lambda_n V^{-1} A^*_{[n]} W_{[n]} \left(b_{[n]} - A_{[n]} x^{(n)} \right). \tag{15}$$

The iteration process from $n = kT$ to $n = (k+1)T$ with Eq. (15) is called one OS cycle. In this paper, we assume that λ_n is constant during one OS cycle, i.e., λ_n for $n = kT + t$ with $t = 0, \ldots, T-1$, are the same; for notation simplicity and without confusion, we will write this constant as λ_k. The general case can also be handled, cf. [16] for more details. When $T = 1$, Eq. (15) reduces to the simultaneous version Eq. (10).

The ART, SART, Cimmino's, DWE and CAV algorithms can be written in the form of Eq. (10) or Eq. (15) with appropriate choices of the diagonal matrices V and W, and the partition set B_t for the ART (the only iterative scheme in the OS format among them), as shown in Table 1.

In this configuration, we can easily derive the OS or simultaneous version of known methods, to generate new algorithms. For example, we can use the partition of the ART and obtain the following OS version of the SART:

$$x_j^{(n+1)} = x_j^{(n)} + \lambda_n \frac{1}{A_{+,j}} \frac{A_{ij}}{A_{i,+}} \left(b^i - A^i x^{(n)} \right) \tag{16}$$

with $i = n \operatorname{mod}(M) + 1$.

Table 1
Choices of V, W and B_t for the ART, SART, Cimmino's and DWE algorithms

ART	$V = I$	$\frac{1}{W^i} = \|A^i\|^2$	$T = M,\ B_t = \{t\}$
SART	$V^j = A_{+,j}$	$\frac{1}{W^i} = A_{i,+}$	
Cimmino's	$V = I$	$\frac{1}{W^i} = M\|A^i\|^2$	
DWE(CAV)	$V = I$	$\frac{1}{W^i} = \sum\limits_{\substack{1 \leqslant l \leqslant N \\ g_{il} \neq 0}} \frac{A_{il}^2}{g_{il}}$	

2.3. Convergence results

In this section, we summarize the convergence results for iterative schemes in the form of the simultaneous Eq. (10) and OS Eq. (15) formats. First, let us state the convergence result specific to the ART, which is instructive to understand the general picture, although most of the results referenced are valid for general schemes. Then, we discuss recent advances.

From a result in [12], if the Eq. (1) is consistent, the sequence $\{x^{(n)}\}$ generated by the ART converges to a solution of Eq. (1) if the relaxation coefficients satisfy

$$0 < \lim_n \inf \lambda_n \leqslant \lim_n \sup \lambda_n < 2. \tag{17}$$

If the Eq. (1) is inconsistent and if the relaxation coefficients satisfy Eq. (17) and are periodic, i.e., $\lambda_n = \lambda_{[n]-1}$, each subsequence, $\{x^{(kM+i)}\}$, $0 \leqslant t < M$, generated by the ART, converges. It was further proved in [6] that if $\lambda_n = \lambda \in (0, 2)$, each sub-sequence $\{x^{(kT+t)}\}$, $0 \leqslant t < M$, generated by the ART, converges to a limit $x_t(\lambda)$ and $\lim_{\lambda\to 0} x_t(\lambda) = x^{(*)} + Px^{(0)}$, where $x^{(*)}$ is the minimal norm solution of Eq. (12), and P is the orthogonal projection from R^N to the kernel of A. The kernel of A is defined as $N(A) = \{x \in R^N : Ax = 0\}$.

For a general OS scheme with T partition sets, $\{x^{(kT+t)}\}$, $0 \leqslant t < T$, are called cyclic sub-sequences. If Eq. (1) is inconsistent, the cyclic sub-sequences converge possibly to T different limits, which consists of the so-called limit cycle, and the intermediate iterated images will oscillate among them. This phenomenon is also termed as semi-convergence [19]. As a result, the reconstructed image may change with different iteration numbers if there exists such a limit cycle. If the images in the limit cycle are far away from each other, the reconstructed image quality will be uncertain, cf. Fig. V.12 in [19]. Because the data are corrupted with noise, Eq. (1) is often inconsistent in practice. Hence, the limit cycle exists with the above relaxation strategy for Eq. (1). Therefore, there is an important and immediate need for an automatic relaxation strategy that removes the limit cycle and guarantees the convergence of the iterative scheme as well as the quality of the reconstructed image.

Recently, Jiang and Wang [16] established the following convergence results for the simultaneous scheme Eq. (10) and the OS scheme Eq. (15). Let $x^{(*)}$ be the minimal solution of Eq. (12) (the V-norm is the induced norm on R^N from the inner product $\langle\cdot,\cdot\rangle_V$ associated with V as the W-norm) and P_V be the orthogonal projection from R^N to the kernel of A with respect to the inner product $\langle\cdot,\cdot\rangle_V$. In the following, $\|A\|_{V,W}$ denotes the operator norm from R^N (with the inner product $\langle\cdot,\cdot\rangle_V$) to R^M (with the inner product $\langle\cdot,\cdot\rangle_W$). Similarly, we can define the operator norm $\|A_t\|_{V,W_t}$. Details can be found in [16]. Their first theorem can be stated as follows:

Theorem 1. (*[16]*) *Assume that there exists* $\rho > 0$ *such that* $\|A_t\|_{V,W_t} \leqslant \rho$ *for* $t = 1, \ldots, T$ *and*

$0 \leqslant \rho^2\lambda_k \leqslant 2$. *If Eq.* (*1*) *is consistent and*

$$\sum_k \min(\rho^2\lambda_k, 2 - \rho^2\lambda_k) = \infty, \tag{18}$$

then the sequence generated by Eq. (*15*) *converges to* $x^{(*)} + P_V x^{(0)}$. *If the partition* (*Eqs* (*13*) *and* (*14*)) *is disjoint and*

$$\sum_k \lambda_k = \infty \quad \text{and} \quad \lim_k \lambda_k = 0, \tag{19}$$

then the sequence generated by Eq. (*15*) *converges to* $x^{(*)} + P_V x^{(0)}$, *even if Eq.* (*1*) *is inconsistent.*

In either case, the sequence converges to a solution of the normal equation Eq. (*12*).

A simple but useful corollary from Theorem 1 is:

Corollary 1. (*[16]*) *Assume that* $\rho = \|A\|_{V,W} > 0$ *and* $0 \leqslant \rho^2\lambda_k \leqslant 2$. *Then, the results of Theorem 1 still hold.*

For the simultaneous scheme Eq. (10), the convergence holds even if $\rho^2\lambda_k \to 2 - 0$.

Theorem 2. *Assume that* $\rho = \|A\|_{V,W} > 0$ *and* $0 \leqslant \rho^2\lambda_k \leqslant 2$. *If Eq.* (*18*) *holds, then the sequence generated by Eq.* (*10*) *converges to* $x^{(*)} + P_V x^{(0)}$, *even if Eq.* (*1*) *is inconsistent.*

The convergence of the ART, SART, Cimmino's Algorithm, DWE and CAV follows from the estimate of $\|A_t\|_{V,W_t}$ or $\|A\|_{V,W}$. With the choices of the matrices V and W, and the partition $\{B_t\}$ for the ART, as shown in Table 1, we can readily show that $\rho = 1$ for all of them [16].

3. EM-like algorithms

If A, x and b are non-negative, the imaging system works in a nonnegative space, where the discrepancy between b and Ax should be quantified using the I-divergence (Kullback-Leibler distance), which is defined as

$$I(b, Ax) = \sum_{i=1}^{M} b^i \log \frac{b^i}{A^i x} + A^i x - b^i. \tag{20}$$

Let $s_j = \sum_{i=1}^{N} A_{ij}$. The EM algorithm [20] is written as

$$x_j^{(n+1)} = x_j^{(n)} \cdot \frac{1}{s_j} \sum_{i=1}^{M} A_{ij} \frac{b^i}{A^i x^{(n)}}. \tag{21}$$

If Eq. (1) is consistent, then the sequence $\{x^{(n)}\}$ generated by Eq. (21) converges to a solution of Eq. (1). If Eq. (1) is inconsistent, then the sequence $\{x^{(n)}\}$ generated by Eq. (21) converges to a minimizer of Eq. (20). This convergence was established in [21].

Let $s_{t,j} = \sum_{i \in B_t} A_{ij}$. The OS version of the EM algorithm, called OSEM, is by replacing both sums in Eq. (21) with partial sums over the partition sets $\{B_t\}$:

$$x_j^{(n+1)} = x_j^{(n)} \cdot \frac{1}{s_{[n],j}} \sum_{i \in B_{[n]}} A_{ij} \frac{b^i}{A^i x^{(n)}}. \tag{22}$$

The convergence of the OSEM algorithm is open in the consistent case, except for the quite special case of subset balance [3,14]. The partition is said to have the subset balance property, if, for each fixed value of j, the sums $s_{i,j}$ are the same for all t, which induces that $s_{t,j} = \frac{s_j}{T}$. If Eq. (1) is inconsistent, there is an example in [2] showing that the OSEM is not convergent.

Let $\omega_t = \max_j \frac{s_{t,j}}{s_j}$. A modified version of the OSEM, called the rescaled block-iterative EM (RBL-EM), was proposed in [3]:

$$x_j^{(n+1)} = \left(1 - \frac{s_{[n],j}}{\omega_{[n]} s_j}\right) x_j^{(n)} + \frac{1}{\omega_{[n]} s_j} x_j^{(n)} \cdot \sum_{i \in B_{[n]}} A_{ij} \frac{b_i}{A^i x^{(n)}}. \tag{23}$$

When the subset balance condition holds, the RBI-EM reduces to the OSEM. The RBI-EM converges, in the consistent case, to a solution of Eq. (1) [3]. In the inconsistent case, the convergence is unknown and it was conjectured that the same semi-convergence similar to that of the ART could happen.

To achieve the convergence in the inconsistent case, the following row-action maximum likelihood algorithm (RAMLA) was discovered by Browne and De Pierro [2]:

$$x_j^{(n+1)} = \left(1 - \lambda_n s_{[n],j}\right) x_j^{(n)} + \lambda_n x_j^{(n)} \cdot \sum_{i \in B_{[n]}} A_{ij} \frac{b_i}{A^i x^{(n)}}. \tag{24}$$

If the relaxation coefficients $\lambda_n > 0$ is constant during one OS run and satisfies the same relaxation strategy Eq. (19), then the sequence $\{x^{(n)}\}$ generated by Eq. (24) converges to a minimizer of Eq. (20), if Eq. (20) is strictly convex in x.

4. Heuristic arguments for the SART and EM

This section is to induce the SART and EM algorithms in a heuristic way. The idea is as follows. An iterative solution to the imaging system Eq. (1) can be an implementation of a feedback scheme, by which a current guess to the solution is refined in the iteration. Specifically, based on a current guess about an underlying image, projection data can be numerically generated. Then, these synthesized data are compared to the observed data according to a discrepancy measure. The most common way to measure a discrepancy between two values is using difference or division. We assume that A and b are non-negative in the following.

4.1. Difference and the SART

Given a current guess $x^{(n)}$, we can synthesize the projection data

$$\overline{b}^i \left(x^{(n)}\right) = \sum_{j=1}^{N} A_{ij} x_j^{(n)}. \tag{25}$$

Then, we compare the observed data b with $\overline{b}$ using the difference $b^i - \overline{b}^i\left(x^{(n)}\right)$ component-wisely. This result is considered as a hint on how to update the current image $x^{(n)}$. There are the following three cases for a given datum b^i:

$$b^i - \overline{b}^i\left(x^{(n)}\right) = \begin{cases} > 0, & \text{i.e., } x^{(n)} \text{ should be somehow increased;} \\ = 0, & \text{i.e., no need to update } x^{(n)}; \\ < 0, & \text{i.e., } x^{(n)} \text{ should be somehow decreased.} \end{cases} \tag{26}$$

For that b_i, we need to rectify $x^{(n)}$ and try to have $b^i = \overline{b}^i\left(x^{(n)}\right)$. The simplest rectification is to add this values to both side of Eq. (34) so that we have

$$\begin{aligned} b^i &= \left[b^i - \overline{b}^i\left(x^{(n)}\right)\right] + \overline{b}^i\left(x^{(n)}\right) = \left[b^i - \overline{b}^i\left(x^{(n)}\right)\right] + \sum_{j=1}^{N} A_{ij} x_j^{(n)} \\ &= \sum_{j=1}^{N} A_{ij}\left[x_j^{(n)} + \frac{b^i - \overline{b}^i\left(x^{(n)}\right)}{A_{i,+}}\right]. \end{aligned}$$

That is, the discrepancy for a given datum b^i can be eliminated by adding $\frac{b^i - \overline{b}^i\left(x^{(n)}\right)}{A_{i,+}}$ to $x_j^{(n)}$. However, the rectifying factor $\frac{b^i - \overline{b}^i\left(x^{(n)}\right)}{A_{i,+}}$ depends on i. Naturally, the overall rectifying factor can be computed by averaging over i according to the relative weight $\frac{A_{ij}}{A_{+,j}}$, which quantifies the relative contribution of $x_j^{(n)}$ to $\overline{b}^i\left(x^{(n)}\right)$. This immediately leads to the SART with relaxation parameter $\lambda_n = 1$:

$$x_j^{(n+1)} = x_j^{(n)} + \sum_{i=1}^{M} \frac{A_{ij}}{A_{+,j}} \frac{b^i - \overline{b}^i\left(x^{(n)}\right)}{A_{i,}}. \tag{27}$$

4.2. Division and the EM

Instead of using difference, we can compare measured data b with $\overline{b}$ using division $\frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}$ component-wisely. For a given datum b^i, this ratio suggests how to modify $x^{(n)}$ according to the following three cases:

$$\frac{b^i}{\overline{b}^i\left(x^{(n)}\right)} = \begin{cases} > 1, & \text{i.e., } x^{(n)} \text{ should be increased;} \\ = 1, & \text{i.e., no need to update } x^{(n)}; \\ < 1, & \text{i.e., } x^{(n)} \text{ should be decreased.} \end{cases} \tag{28}$$

For that b_i and the known $\frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}$, we can multiply both sides of Eq. (34) with the ratio for a perfect rectification: $b^i = \frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}\overline{b}^i\left(x^{(n)}\right) = \frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}\sum_{j=1}^{N} A_{ij} x_j^{(n)}$

$$= \sum_{j=1}^{N} A_{ij}\left[x_j^{(n)} \frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}\right].$$

However, the rectifying factor $\frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}$ depends on i. Similarly, the overall rectifying factor can be computed by averaging over i according to the relative weight $\frac{A_{ij}}{A_{+,j}}$. Then, we have the EM formula:

$$x_j^{(n+1)} = x_j^{(n)} \cdot \sum_{i=1}^{M} \frac{A_{ij}}{A_{+,j}} \frac{b^i}{\overline{b}^i\left(x^{(n)}\right)}. \tag{29}$$

5. Discussions and conclusion

The investigation on the OS methods is a current hot topic in the iterative reconstruction field. The ART is the first OS iterative scheme used in CT, which initialized the studies on various row-action methods [5]. The Cimmino's algorithm can be written as the simultaneous version of the ART up to a rescaling of the relaxation coefficient. By the theorems in Section 2.3, the permissible interval of the relaxation coefficients is [0,2] for the ART, but the equivalent interval of the Cimmino's algorithm is $\left[0, \frac{2}{M}\right]$, much smaller that that of the ART. That may induce a larger update per iteration, and is why the OS schemes are often preferred. However, major efforts are needed for further evaluation and comparison.

Given various algorithms, one needs to know which one to use for a specific application. There are pros and cons with each method [9], hence the specifics of the problem and requirements by users must be matched to the capabilities of an algorithm. Some guidelines may be given as follows. First, one may consider which space the problem is in and what objective functional should be used. In the real/complex space, the Landweber scheme can be the primary choice. In the nonnegative space, the EM-like approach would be preferred. It has been found in recent experiments that the OS algorithms can produce satisfactory image quality with an order-of-magnitude fewer iterations than its simultaneous version [14]. To perform OS iterations, the relaxation strategies discussed in this review should be followed so that the convergence is approached.

The results in Theorem 1, Corollary 1 and Theorem 2 specify the conditions for the relaxation strategy to remove the limit cycle for a general class of algebraic algorithms. In addition, these results characterize the structure of the final limit, and reveal the dependence of the final limit on the initial image. One important corollary from those results is that the final limit is continuous with respect to the initial guess. Note that this is not a trivial result; nothing like this has been known for the EM-like algorithms. Some results, similar to but weaker than that in Theorem 1, were obtained by Trummer for the ART [22,23]. The convergence or semi-convergence of the SART was proved under various relaxation conditions [4, 7,15]. However, the result in this review, when applied to the SART, surpasses all those results.

Despite the recent advances, there remain numerous problems for further research. One widely used approach in image reconstruction is to use regularization in terms of prior. How to optimally integrate prior knowledge remains a primary issue. Other topics include selection of the initial image, specification of the block size (the size of the ordered subsets), optimization of the relaxation strategy, optimal termination of the iterative process, and so on. Furthermore, the iterative schemes can be studied under certain constraints, such as nonnegative and linear constraints.

In conclusion, iterative algorithms for image reconstruction have been reviewed, with an emphasis on the EM-like and ART-like algorithms in both of their simultaneous and order-subset formats. Major theoretical and practical issues have been discussed for the understanding and applications of these algorithms. Additionally, research directions have been outlined. Iterative algorithms would be more important in the future with the rapid development of the computer technology.

Acknowledgement

The first author performed this work while he was a visiting professor with CT/Micro-CT Laboratory, Department of Radiology, University of Iowa. His work was supported in part by a Chinese NKBRSF grant (G1998030606) and an NIH grant (R01 DC03590). The second author was supported in part by an NIH grant (R01 DC03590).

References

[1] A.H. Andersen and A.C. Kak, Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm, *Ultrasonic Imaging* **6** (1984), 81–94.
[2] J. Browne and A.R. De Pierro, A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography, *IEEE Trans. Medical Imaging* **15**(5) (1996), 687–699.
[3] C.L. Byrne, *Notes on the EMML algorithm and related block-iterative methods*, Technical report, Department of Mathematical Sciences, University of Massachusetts, Lowell, 2001.
[4] C.L. Byrne and J. Graham-Eagle, *A short proof of the convergence of the SART algorithm*, Technical report, Department of Mathematical Sciences, University of Massachusetts, Lowell, 2001.
[5] Y. Censor, Row-action methods for huge and sparse systems and their applications, *SIAM Review* **23**(4) (1981), 444–466.
[6] Y. Censor, P.P.B. Eggermont and D. Gordon, Strong underrelaxation in Kaczmarz's method for inconsistent systems, *Numerische Mathematik* **41** (1983), 83–92.
[7] Y. Censor and T. Elfving, Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem, *SIAM Journal on Matrix Analysis and Applications* (to appear).
[8] Y. Censor, D. Gordon and R. Gordon, Component averaging: an efficient iterative parallel algorithm for large and sparse unstructured problems, *Parallel Computing* **27** (2001), 777–808.
[9] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, 1997.
[10] G. Cimmino, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*, La Ricerca Scientifica XVI, Series II, Anno IX, 1938, pp. 326–333.
[11] A. Dax, The convergence of linear stationary iterative processes for solving singular unstructured systems of linear equations, *SIAM Review* **32**(4) (1990), 611–635.
[12] P.. B. Eggermont, G.T. Herman and A. Lent, Iterative algorithms for larger partitioned linear systems, with applications to image reconstruction, *Linear Algebra and Its Applications* **40** (1981), 37–67.
[13] R. Gordon, R. Bender and G.T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography, *Journal of Theoretical Biology* **29** (1970), 471–482.
[14] H.M. Hudson and R.S. Larkin, Accelerated image reconstruction using ordered subsets of projection data, *IEEE Trans. Medical Imaging* **13**(4) (1994), 601–609.
[15] M. Jiang and G. Wang, Convergence of the simultaneous algebraic reconstruction technique (SART), in *Proceedings of the 35th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Invited talk, 2001.
[16] M. Jiang and G. Wang, *Convergence Studies on Iterative algorithms for image reconstruction*, Technical report, Micro-CT Laboratory, Department of Radiology, University of Iowa, Iowa City, IA, 2002.
[17] S. Kaczmarz, Angen herte aufl sung von systemn linearer gleichungen, *Bull. Acad. Polon. Sci. Lett.* **A35** (1937), 335–357.
[18] R. Leahy and C.L. Byrne, Recent developments in iterative image reconstruction for PET and SPECT, *IEEE Trans. Medical Imaging* **19** (2001), 257–260.
[19] Natterer, *The Mathematics of Computerized Tomography*, John Wiley & Sons, New York, 1986.
[20] L.A. Shepp and Y. Vardi, Maximum likelihood restoration for emission tomography, *IEEE Trans. Med. Imaging* **1** (1982), 113–122.
[21] D.L. Snyder, T.J. Schulz and J.A. O'Sullivan, Deblurring subject to nonnegativity constraints, *IEEE Trans. Signal Processing* **40** (1992), 1143–1150.
[22] M.R. Trummer, Reconstructing pictures from projections: on the convergence of the ART algorithm with relaxation, *Computing* **26**(3) (1981), 189–195.
[23] M.R. Trummer, A note on the ART of relaxation, *Computing* **3**(3–4) (1984), 349–352.