To appear in International Journal of Computer Mathematics

Convergence Analysis of SART: Optimization and Statistics

Ming Yan

Simultaneous algebraic reconstruction technique (SART) [1, 2] is an iterative method for solving inverse problems of form $Ax(+n) = b$. This type of problems arises for example in computed tomography reconstruction, in which case A is obtained from discrete Radon transform. In this paper, we provide several methods for derivation of SART and connections between SART and other methods. Using these connections, we also prove the convergence of SART in different ways. These approaches are from optimization and statistical points of view and can be applied to other Landweber-like schemes such as Cimmino's algorithm and component averaging (CAV). Furthermore, the noisy case is considered and error estimation is given. Several numerical experiments for computed tomography reconstruction are provided to demonstrate the convergence results in practice.

Keywords. simultaneous algebraic reconstruction technique, Bregman iteration, dual gradient descent, expectation maximization, alternating minimization, image reconstruction, component averaging, Cimmino's algorithm.

1 Introduction

As a group of methods for reconstructing two dimensional and three dimensional images from projections of objects, iterative reconstruction has many applications in such as computed tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI). This technique is quite different from filtered back projection (FBP) [25], which is the most commonly used algorithm in practice by manufacturers [22]. The main advantages of the iterative reconstruction technique over FBP are insensitivity to noise and flexibility [16]. The data can be collected over any set of lines; the projections do not have to be distributed uniformly in angle and may be even incomplete.

The inverse problem to be solved is based on the system of linear equations

$$Ax(+n) = b,$$

where $x = (x_1, \cdots, x_N)^T \in \mathbf{R}^N$ is the unknown image to be reconstructed from the projections and expressed as a long vector. N depends on the resolution of the image to be reconstructed. If the image to be reconstructed is

Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095.

256x256, then $N = 256 \times 256$. b contains all the given (noisy) measurements with $b = (b_1, \cdots, b_M)^T \in \mathbf{R}^M$. M is the number of measurements of the image. A is a $M \times N$ matrix describing the transformation from x to the measurements. In computed tomography, A represents discrete Radon transform using Siddon's algorithm [27], with each row describing an integral along one straight line, and all the elements are nonnegative. If there is no noise in the measurements ($Ax = b$), b is in the range of A and at least one solution exists. Otherwise, if there is noise in the measurements, we have $Ax + n = b$ with n being the unknown noise and b may not be in the range of A. If M is large enough to make the matrix having full column rank, the solution is unique for the noise free case. Otherwise, if the measurements are not sufficient enough to make the matrix A having full column rank, there will be infinite many solutions for the noise free case, and an iterative reconstruction algorithm will converge to one solution, which will usually depend on the initial guess.

There are many available iterative reconstruction algorithms to find a solution of the inverse problem and some examples of iterative reconstruction algorithms are expectation maximization (EM) [26], algebraic reconstruction techniques (ART) [11, 13], and component averaging methods (CAV) [7]. Simultaneous algebraic reconstruction technique (SART) [1, 2], as a refinement of ART, is also widely used [3, 23, 33].

The focus of this paper is SART. In fact the convergence analysis of SART and CAV is studied by Wang and Zheng [29], Censor and Elfving [7], Jiang and Wang et. al. [14, 15, 24]. We will show several new derivations of SART from optimization and statistical points of view: I) SART will be derived from linearized Bregman iteration for a primal optimization problem and gradient descent with unit step for the corresponding dual problem; II) The connection between SART and EM is provided by showing that SART is an EM algorithm and alternating minimization method. From the connections between SART and these methods, we have many different ways to show the convergence for SART.

The organization of this paper is as follows: in section 2, we will give an introduction to SART. The concept of Bregman iteration and linearized Bregman iteration, along with the connection of SART and linearized Bregman iteration of a primal optimization problem, will be provided in section 3. Then we will show the convergence analysis of SART using linearized Bregman iteration and dual gradient descent in sections 4 and 5 respectively. A connection between SART and expectation maximization algorithm is provided in section 6 to derive SART from both statistical and optimization points of view. Also, we show that the convergence analysis can be applied to a general Landweber-like scheme [12,17,18]. Furthermore, we discuss the noisy case in section 7. In order to illustrate the convergence results shown in this paper, we present numerical experiments for both noise free and noisy cases in section 8, corresponding to computed tomography reconstruction. Finally, we end up with a conclusion section.

2 Simultaneous Algebraic Reconstruction Technique

In this section, we provide an introduction to simultaneous algebraic reconstruction technique (SART) [1, 2]. The system of linear equations mentioned in section 1 is $Ax(+n) = b$, where all elements in A are nonnegative (this is a property that must hold when A is obtained from discrete Radon transform in computed tomography reconstruction). The objective of the reconstruction technique is to find a solution x of this system. In this section, we only consider the noise free case ($Ax = b$), and the existence of solutions is guaranteed, the noisy case where b is not in the range of A is considered in section 7. However, the uniqueness is still unknown. If A has full column rank, the solution to the system is unique. Otherwise, there are infinite many solutions to this system. SART is a method used to find one solution, which depends on the initial guess.

We define two diagonal matrices V and W as follows:

$$V = \mathrm{diag}(\mathbf{1}^T A), \qquad W = \mathrm{diag}(A\mathbf{1}),$$

where $\mathbf{1}$ is the column vector in $\mathbf{R}^M$ (or $\mathbf{R}^N$) with all entries equal to one. In addition, we assume that $W_{i,i} > 0$ and $V_{j,j} > 0$. Actually if $W_{i,i} = 0$, the i^{th} measurement b_i is 0 for all x. $V_{j,j} = 0$ means that changes in the j^{th} component of x cannot be detected in the measurements. Then SART proposed in [1, 2] is

$$x^{k+1} = x^k + wV^{-1}A^T W^{-1}(b - Ax^k), \tag{1}$$

for $k = 0, 1, \cdots$, where w is a relaxation parameter in $(0, 2)$, and the starting point is x^0. Here, x^0 can be $\mathbf{0}$ (black image as a column vector with all entries equal to zero) or output images from other methods (e.g. FBP).

3 SART from Bregman Iteration

Before deriving SART from Bregman iteration, we provide some definitions for later use. First we recall the definition of ellipsoidal norms for vectors. Let G be a $K \times K$ symmetric positive definite matrix. The ellipsoidal norm of

$x \in \mathbf{R}^K$ with respect to G is defined as follows:

$$\|x\|_G^2 = \langle x, x\rangle_G = \langle x, Gx\rangle = x^T Gx.$$

Since different matrices imply different ellipsoidal norms, we indicate the positive definite matrices in the notation of these norms. The ellipsoidal norm with matrix G is named as G-norm. Two ellipsoidal norms $\|\cdot\|_V$ and $\|\cdot\|_{W^{-1}}$, named V-norm and W^{-1}-norm respectively, will be used.

From the convergence analysis of SART in [14], SART converges to the solution of $Ax = b$ with the least V-norm if the initial guess x^0 is $\mathbf{0}$. We consider the following primal optimization problem which depends on the initial guess x^0,

$$\begin{cases} \underset{x}{\text{minimize}} & \|x - x^0\|_V^2, \\ \text{subject to} & Ax = b. \end{cases} \tag{2}$$

This is a convex constrained optimization problem with one unique solution. We can approximate this problem by adding a quadratic penalty function of the equality constraint onto the objective function to obtain an unconstrained problem as follows,

$$\underset{x}{\text{minimize}}\, \mu\|x - x^0\|_V^2 + \frac{1}{2}\|Ax - b\|_{W^{-1}}^2, \tag{3}$$

where μ is a positive parameter.

The solution of (3) converges to the solution of (2) as $\mu \to 0$ [19]. Thus, we can find an approximate solution of (2) by solving (3) with a sufficient small μ. However, small μ will slow down many algorithms and may even lead to ill-conditioning [19].

For solving constrained problems, Bregman proposed a method by solving a sequence of unconstrained problems [5]. This iterative method is recently used in compressive sensing and image processing for solving ℓ_1 minimization problems [20]. In addition, a linearized version of Bregman iteration, named linearized Bregman iteration, is proposed by Osher et al. [21] and Yin et al. [32].

Instead of solving one unconstrained problem to obtain an approximate solution of problem (2), we solve several unconstrained problems iteratively, and the result also converges to the solution of the constrained problem (2).

For any convex function $J(x)$, the Bregman distance is defined by

$$D_J^p(x^1, x^2) = J(x^1) - J(x^2) - \langle p, x^1 - x^2\rangle,$$

for all x^1 and x^2 in the domain of J. Here $p \in \partial J(x^2)$ is a subgradient of J at x^2 if J is not differentiable, which means $J(x) - J(x^2) \geq \langle p, x - x^2 \rangle$ for any x in the domain of J. When J is differentiable, p is the gradient of J at x^2. For our case, J is defined as $J(x) = \mu \|x - x^0\|_V^2$, which is differentiable, and we have $D_J^p(x^1, x^2) = \mu \|x^1 - x^2\|_V^2$.

The Bregman iteration is as follows:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}} \; D_J^p(x, x^k) + \frac{1}{2}\|Ax - b\|_{W^{-1}}^2 \\ &= \underset{x}{\operatorname{argmin}} \; \mu\|x - x^k\|_V^2 + \frac{1}{2}\|Ax - b\|_{W^{-1}}^2, \end{aligned} \tag{4}$$

for $k = 0, 1, \cdots$, starting with initial guess x^0. In each iteration, we can find the optimal solution analytically. From the optimality of x^{k+1} in (4), it follows that

$$2\mu V(x^{k+1} - x^k) + A^T W^{-1}(Ax^{k+1} - b) = \mathbf{0}.$$

Therefore, for each iteration, we have to solve a system of linear equations

$$(2\mu V + A^T W^{-1} A)x^{k+1} = A^T W^{-1} b + 2\mu V x^k. \tag{5}$$

The matrix on the left-hand-side remains unchanged while the right-hand-side changes for each iteration. If N is small, it is easy to solve for x^{k+1} by precalculating the inverse matrix. While for large N, direct methods would be prohibitively expensive and may even impossible in some cases. Iterative methods are needed to solve for x^{k+1}. To avoid solving many large systems of linear equations, we can linearize the second quadratic term in (4) and obtain the linearized Bregman iteration as follows,

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \; \mu\|x - x^k\|_V^2 + \langle Ax^k - b, Ax \rangle_{W^{-1}} + \frac{1}{2\alpha}\|x - x^k\|_V^2, \tag{6}$$

where $\alpha \in (0, 2/\rho(V^{-\frac{1}{2}} A^T W^{-1} A V^{-\frac{1}{2}}))$ is chosen to make it converge. In this linearized version, the optimal solution is also easy to find, and it is given by

$$x^{k+1} = x^k + \frac{1}{2\mu + \frac{1}{\alpha}} V^{-1} A^T W^{-1}(b - Ax^k).$$

Denoting $w = 1/(2\mu + \frac{1}{\alpha})$, we thus observe that SART is now derived from the linearized Bregman iteration. Then we can find the convergence of SART using the linearized Bregman method, as presented next.

4 Convergence Analysis of SART

The convergence of linearized Bregman iteration has been shown for different problems in compressive sensing and image processing [6, 31]. In this section, we will show the convergence of SART from linearized Bregman iteration, using some special properties of SART. First we state the following important lemma [15], which will be used for several times.

Lemma 4.1 *Matrices $V - A^T W^{-1} A$ and $W - AV^{-1}A^T$ are positive semidefinite.*

Remark 1 From lemma 4.1, we have $x^T V x - x^T A^T W^{-1} A x \geq 0$ for all $x \in \mathbf{R}^N$, which means $\|Ax\|_{W^{-1}} \leq \|x\|_V$. In fact, $V - A^T W^{-1} A$ and $W - AV^{-1}A^T$ are not positive definite, because $\mathbf{1}^T V \mathbf{1} - \mathbf{1}^T A^T W^{-1} A \mathbf{1} = 0$ and $\mathbf{1}^T W \mathbf{1} - \mathbf{1}^T A V^{-1} A^T \mathbf{1} = 0$.

At first, we will show that the residual $b - Ax^k$ is decreasing with respect to the W^{-1}-norm by the following theorem with a proof different from [14].

Theorem 4.2 *For x^k obtained by (1), the W^{-1}-norm of the residual $b - Ax^k$ is decreasing if $0 < w < 2$. Furthermore, we have the following inequality:*

$$\|Ax^{k+1} - b\|^2_{W^{-1}} + (\frac{2}{w} - 1)\|x^{k+1} - x^k\|^2_V \leq \|Ax^k - b\|^2_{W^{-1}}, \tag{7}$$

for $k = 0, 1, \cdots$.

Proof From the iteration $x^{k+1} = x^k + wV^{-1}A^T W^{-1}(b - Ax^k)$, we have

$$\begin{aligned}
&\langle Ax^k - b, A(x^{k+1} - x^k)\rangle_{W^{-1}} \\
&= \langle W^{-1}(Ax^k - b), A(x^{k+1} - x^k)\rangle = \langle A^T W^{-1}(Ax^k - b), (x^{k+1} - x^k)\rangle \\
&= -\langle \frac{1}{w} V(x^{k+1} - x^k), (x^{k+1} - x^k)\rangle = -\frac{1}{w}\|x^{k+1} - x^k\|^2_V,
\end{aligned}$$

which is equivalent to

$$\frac{1}{w}\|x^{k+1} - x^k\|^2_V + \langle Ax^k - b, Ax^{k+1} - Ax^k\rangle_{W^{-1}} = 0. \tag{8}$$

In addition, we have

$$
\begin{aligned}
&2\langle Ax^k - b, Ax^{k+1} - Ax^k\rangle_{W^{-1}} \\
=&\langle Ax^{k+1} - b, Ax^{k+1} - b\rangle_{W^{-1}} - \langle Ax^{k+1} - Ax^k, Ax^{k+1} - Ax^k\rangle_{W^{-1}} \\
&- \langle Ax^k - b, Ax^k - b\rangle_{W^{-1}}.
\end{aligned}
$$

Plugging this into equation (8) implies that

$$
\frac{2}{w}\|x^{k+1} - x^k\|_V^2 - \|Ax^{k+1} - Ax^k\|_{W^{-1}}^2 + \|Ax^{k+1} - b\|_{W^{-1}}^2 = \|Ax^k - b\|_{W^{-1}}^2.
$$

The remark after Lemma 4.1 gives $\|Ax^{k+1} - Ax^k\|_{W^{-1}}^2 \le \|x^{k+1} - x^k\|_V^2$.

Thus, we obtain

$$
(\frac{2}{w} - 1)\|x^{k+1} - x^k\|_V^2 + \|Ax^{k+1} - b\|_{W^{-1}}^2 \le \|Ax^k - b\|_{W^{-1}}^2.
$$

If $0 < w < 2$, the residual is decreasing in W^{-1}-norm. □

From the above theorem, the residual will decrease in the W^{-1}-norm until x^k remains unchanged. We will show that the sequence $\{x^k\}$ will converge and converges to the solution of the constrained problem (2). The convergence of $\{x^k\}$ is shown firstly. Since

$$
x^{k+1} - x^k = (I - wV^{-1}A^TW^{-1}A)(x^k - x^{k-1}),
$$

denoting $r^k = x^{k+1} - x^k$ gives us

$$
\begin{aligned}
V^{\frac{1}{2}}r^k &= V^{\frac{1}{2}}(I - wV^{-1}A^TW^{-1}A)V^{-\frac{1}{2}}V^{\frac{1}{2}}r^{k-1} \\
&= (I - wV^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}})V^{\frac{1}{2}}r^{k-1}. \qquad (9)
\end{aligned}
$$

The exponential decay of $\|r^k\|_V$ or $\|V^{\frac{1}{2}}r^k\|$ is shown by proving

$$
0 \prec wV^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}} \prec 2I \qquad (10)
$$

during the iterations. For two symmetric positive semidefinite matrices A and B, $A \prec (\preceq)B$ means that $B - A$ are positive definite (positive semidefinite). Lemma 4.1 implies $A^TW^{-1}A \preceq V$, which is equivalent to

$$
V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}} \preceq I.
$$

Thus for $w \in (0,2)$, we have

$$wV^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}} \prec 2I.$$

However, $0 \prec V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}$ is not always true. If A has nontrivial null space $\mathcal{N}(A)$, we can choose a nontrivial x such that $V^{-\frac{1}{2}}x \in \mathcal{N}(A)$ and

$$x^TV^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}x = 0.$$

But in a subspace of R^N, the image of $V^{-\frac{1}{2}}A^T$, we have $0 \prec V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}$.

From the iteration $x^{k+1} = x^k + wV^{-1}A^TW^{-1}(b - Ax^k)$, we have

$$V^{\frac{1}{2}}r^k = V^{\frac{1}{2}}(x^{k+1} - x^k) = wV^{-\frac{1}{2}}A^TW^{-1}(b - Ax^k).$$

Thus $V^{\frac{1}{2}}r^k$ is in the image of $V^{-\frac{1}{2}}A^T$, and $0 \prec V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}$ is satisfied during the iterations.

Therefore, for $w \in (0,2)$, from the exponential decay of $\|r^k\|_V$ in (9) and (10), there exists a constant $\lambda \in (0,1)$ such that $\|x^{k+1} - x^k\|_V \le \lambda\|x^k - x^{k-1}\|_V \le \lambda^k\|x^1 - x^0\|_V$. Assuming that x^k converges to $\bar{x}$ and we have

$$\|x^k - \overline{x}\|_V \le \sum_{i=k}^{\infty}\|x^{i+1} - x^i\|_V \le \sum_{i=k}^{\infty}\lambda^i\|x^1 - x^0\|_V = \frac{\lambda^k}{1-\lambda}\|x^1 - x^0\|_V. \quad (11)$$

We will show next that $\bar{x}$ is the solution of the constrained problem (2).

Lemma 4.3 *Assuming that x^* is the solution of the constrained problem (2) and x^k converges to $\bar{x}$, we have the following estimate:*

$$\mu\|\overline{x} - x^0\|_V^2 \le \mu\|x^* - x^0\|_V^2 + \frac{1}{\alpha}\langle\overline{x} - x^0, x^* - \overline{x}\rangle_V. \quad (12)$$

Proof From the updating of x^k, we have

$$\begin{aligned} x^k &= x^{k-1} + wV^{-1}A^TW^{-1}(b - Ax^{k-1}) \\ &= x^{k-2} + wV^{-1}A^TW^{-1}(b - Ax^{k-2}) + wV^{-1}A^TW^{-1}(b - Ax^{k-1}) \\ &= \cdots = x^0 + wV^{-1}A^TW^{-1}\sum_{i=0}^{k-1}(b - Ax^i). \end{aligned} \quad (13)$$

After substituting $w = 1/(2\mu + \frac{1}{\alpha})$ into (13), we can obtain

$$2\mu(x^k - x^0) = V^{-1}A^TW^{-1}\sum_{i=0}^{k-1}(b - Ax^i) - \frac{1}{\alpha}(x^k - x^0).$$

The nonnegativity of Bregman distance implies that

$$\begin{aligned}
&\mu\|x^k - x^0\|_V^2 \\
&\leq \mu\|x^* - x^0\|_V^2 - 2\mu\langle x^* - x^k, x^k - x^0\rangle_V \\
&= \mu\|x^* - x^0\|_V^2 - \langle x^* - x^k, V^{-1}A^TW^{-1}\sum_{i=0}^{k-1}(b - Ax^i) + \frac{1}{\alpha}x^k - x^0\rangle_V \\
&= \mu\|x^* - x^0\|_V^2 - \langle b - Ax^k, \sum_{i=0}^{k-1}(b - Ax^i)\rangle_{W^{-1}} + \frac{1}{\alpha}\langle x^* - x^k, x^k - x^0\rangle_V.
\end{aligned}$$

We will show that $\|Ax^k - b\|_{W^{-1}}$ decays exponentially. Then the middle term will vanish when $k \to \infty$ and the result follows.

Recall the iteration $x^{k+1} = x^k + wV^{-1}A^TW^{-1}(b - Ax^k)$. Multiplying by A and subtracting b for both sides, we have

$$\begin{aligned}
Ax^{k+1} - b &= Ax^k - b - wAV^{-1}A^TW^{-1}(Ax^k - b) \\
&= (I - wAV^{-1}A^TW^{-1})(Ax^k - b),
\end{aligned}$$

and similarly to the proof of exponential decreasing of $\|r^k\|_V$, we can show exponential decreasing of $\|Ax^k - b\|_{W^{-1}}$ for $w \in (0, 2)$. □

THEOREM 4.4 *Given the same assumption as lemma 4.3, x^k converges to x^*, i.e., $\overline{x} = x^*$.*

Proof From the proof of the previous lemma, residual $b - Ax^k$ decreases exponentially in W^{-1}-norm. Thus $A\overline{x} = b$ since $x^k \to \overline{x}$. In addition, from the assumption that $Ax^* = b$, we have $A(\overline{x} - x^*) = \mathbf{0}$. Furthermore, $\overline{x} = x^0 + wV^{-1}A^TW^{-1}\sum_{i=0}^{\infty}(b - Ax^i)$, which means that $\overline{x} - x^0$ is in the image of $V^{-1}A^T$, and $\langle \overline{x} - x^0, x^* - \overline{x}\rangle_V = 0$. Thus the last term in (12) vanishes and $\|\overline{x} - x^0\|_V^2 \leq \|x^* - x^0\|_V^2$. It follows that $\overline{x} = x^*$ because x^* is the solution of $Ax = b$ with the smallest $\|x - x^0\|_V$. □

From the convergence proof, we can see the importance of the initial guess. If the matrix A does not have full column rank, SART will converge to a solution of the system having the shortest distance to the initial guess x^0 in V-norm. This will be also seen in numerical experiments.

5 Dual Gradient Descent

The connection between linearized Bregman iteration and gradient descent with unit step for dual problem is shown in [31] by Yin for compressive sensing problems. Here we can also derive the convergence of SART by considering the gradient descent method of the dual problem. Assume that the primal optimization problem is

$$\begin{cases} \underset{x}{\text{minimize}} & \frac{1}{2w}\|x-x^0\|_V^2, \\ \text{subject to} & W^{-\frac{1}{2}}Ax = W^{-\frac{1}{2}}b, \end{cases}$$

which is equivalent to the constrained problem (2).

The Lagrangian function [4] of the primal problem is

$$L(x,y) = \frac{1}{2w}\|x-x^0\|_V^2 + y^T W^{-\frac{1}{2}}(Ax-b),$$

with y being the Lagrangian coefficients corresponding to the equality constraints $W^{-\frac{1}{2}}Ax = W^{-\frac{1}{2}}b$. Minimizing $L(x,y)$ with respect to x only provides the optimal solution

$$x = x^0 - wV^{-1}A^T W^{-\frac{1}{2}}y.$$

Plugging this into the Lagrangian function, we have

$$\begin{aligned} \min_x L(x,y) &= \frac{w}{2}\|V^{-1}A^T W^{-\frac{1}{2}}y\|_V^2 - y^T W^{-\frac{1}{2}}(wAV^{-1}A^T W^{-\frac{1}{2}}y + b - Ax^0) \\ &= -y^T W^{-\frac{1}{2}}(b-Ax^0) - \frac{w}{2}\|V^{-1}A^T W^{-\frac{1}{2}}y\|_V^2. \end{aligned}$$

Therefore, the dual problem is

$$\underset{y}{\text{minimize}}\, F(y) = y^T W^{-\frac{1}{2}}(b-Ax^0) + \frac{w}{2}\|V^{-1}A^T W^{-\frac{1}{2}}y\|_V^2, \tag{14}$$

and this is an unconstrained problem, which can be solved by many methods for unconstrained optimization problems.

First of all, we look at the gradient descent method for this problem and see the connection with SART. The gradient of $F(y)$ is

$$\nabla F(y) = W^{-\frac{1}{2}}(b - Ax^0) + wW^{-\frac{1}{2}}AV^{-1}A^TW^{-\frac{1}{2}}y,$$

and the gradient descent method with unit step is

$$y^{k+1} - y^k = -\nabla F(y^k) = -W^{-\frac{1}{2}}(b - Ax^0) - wW^{-\frac{1}{2}}AV^{-1}A^TW^{-\frac{1}{2}}y^k.$$

Multiply by $-wV^{-1}A^TW^{-\frac{1}{2}}$, and we have

$$\begin{aligned}(x^{k+1} - x^0) - (x^k - x^0) &= wV^{-1}A^TW^{-1}(b - Ax^0) - wV^{-1}A^TW^{-1}A(x^k - x^0)\\ &= wV^{-1}A^TW^{-1}(b - Ax^k),\end{aligned}$$

which is again exactly SART. Therefore, SART is equivalent to the gradient descent method with unit step for the dual problem.

The convergence can be derived from gradient descent. Since for any two vectors y^1 and y^2 in $\mathbf{R}^M$, we have

$$\|\nabla F(y^1) - \nabla F(y^2)\| = w\|W^{-\frac{1}{2}}AV^{-1}A^TW^{-\frac{1}{2}}(y^1 - y^2)\| \leq w\|y^1 - y^2\|.$$

Therefore

$$\begin{aligned}F(y^{k+1}) &\leq F(y^k) + \langle y^{k+1} - y^k, \nabla F(y^k)\rangle + \frac{w}{2}\|y^{k+1} - y^k\|^2\\ &= F(y^k) + \left(\frac{w}{2} - 1\right)\|y^{k+1} - y^k\|^2.\end{aligned}$$

$F(y^k)$ will decrease until y^k remains unchanged. In addition, we have

$$y^{k+1} - y^k = (1 - wW^{-\frac{1}{2}}AV^{-1}A^TW^{-\frac{1}{2}})(y^k - y^{k-1}).$$

Similarly, we can show the exponential decay of $\|y^{k+1} - y^k\|$ and $\|y^k - y^*\|$ with y^* being the optimal solution, therefore the exponential decay of $\|x^{k+1} - x^k\|_V$ and $\|x^k - x^*\|_V$, since

$$\begin{aligned}\|x^{k+1} - x^k\|_V &= w\|V^{-\frac{1}{2}}A^TW^{-\frac{1}{2}}(y^{k+1} - y^k)\| = w\|A^TW^{-\frac{1}{2}}(y^{k+1} - y^k)\|_{V^{-1}}\\ &\leq w\|W^{-\frac{1}{2}}(y^{k+1} - y^k)\|_W = \|y^{k+1} - y^k\|.\end{aligned}$$

Thus, the gradient descent is convergent given the step size (here is 1) is less than $2/w$, which is $1 < 2/w$ or $w < 2$.

6 SART is an Expectation Maximization Algorithm

Among all the iterative algorithms for CT reconstruction, there are two important classes. One is algebraic reconstruction technique (ART) [11, 13] based methods. Simultaneous algebraic reconstruction technique (SART) [1, 2], as a refinement of ART, is used widely. The other is expectation maximization (EM) [10, 26] from statistical model. The EM algorithm is a general approach for maximizing *a posterior* distribution when some of the data is missing. It is an iterative method which alternates between expectation (E) steps and maximization (M) steps. These two classes are considered as two different groups. In this section, we will show that SART is also an EM algorithm, building up the connection between these two classes. We will derive SART using EM by introducing latent variables, which is assumed to be missing.

In fact, the sequence generated by SART converges to a minimizer of the weighted least square function $\|b-Ax\|_{W^{-1}}$ from any initial guess [14]. Therefore, SART can be considered as an algorithm for solving the following maximum likelihood problem:

$$\operatorname*{maximize}_{x} p_B(b|x) = \prod_i \frac{1}{\sqrt{2\pi W_{i,i}}} e^{-\frac{(b_i-(Ax)_i)^2}{2W_{i,i}}}. \tag{15}$$

b_i is assumed to follow normal distribution with expectation $(Ax)_i$ and variance $W_{i,i}$, where the variance of each measurement is the length of the intersection of X-ray from source to detector with the image domain.

In order to show that SART is EM, we introduce latent variables $\{z_{ij}\}$ as the measurements of pixel j by detector i and assume that they follow normal distributions with expected values $\{A_{i,j}x_j\}$ and variances $\{A_{i,j}\}$ respectively, because the summation of two normal distributed random variables also follows a normal distribution with expected value being summation of the two expected values and variance being summation of the two variances. The original E-step is to find the expectation of the log-likelihood given the present variables x^k and the constraints $b_i = \sum_j z_{ij}$. It is easy to derive that, under the constraints, $\{z_{ij}\}$ are still realizations of normally distributed random variables with expected values $\{A_{i,j}x_j + \frac{A_{i,j}(b_i-(Ax)_i)}{W_{i,i}}\}$ and variances $\{\frac{A_{i,j}(W_{i,i}-A_{i,j})}{W_{i,i}}\}$ respectively.

The M-step is to maximize the expected value of the log-likelihood function,

$$\begin{aligned} E_{z|x^k,b}\log p(b,z|x) &= -E_{z|x^k,b}\sum_{ij}\frac{(z_{ij}-A_{i,j}x_j)^2}{2A_{i,j}} + \text{constant} \\ &= -\sum_{ij}\frac{(E_{z|x^k,b}z_{ij}-A_{i,j}x_j)^2}{2A_{i,j}} + \text{constant}. \end{aligned}$$

Therefore, for E-step we have to just find the expected value of z_{ij} given x^k and the constraints, which is

$$z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{A_{i,j}(b_i-(Ax^k)_i)}{W_{i,i}}. \tag{16}$$

After $\{z_{ij}\}$ are updated, x^{k+1} is obtained by maximizing $p(b, z^{k+1}|x)$, which has an analytical solution:

$$x_j^{k+1} = \frac{\sum_i z_{ij}^{k+1}}{V_{j,j}} = x_j^k + \frac{1}{V_{j,j}}\sum_i \frac{A_{i,j}(b_i-(Ax^k)_i)}{W_{i,i}}. \tag{17}$$

This is the original SART algorithm proposed by Andersen [1].

6.1 *SART is an Alternating Minimization Method*

With the latent variables $\{z_{ij}\}$, SART can also be derived from an alternating minimization method of another problem with variables x and z. The new problem is

$$\begin{cases} \underset{x,z}{\text{minimize}}\ E(x,z) := \sum_{ij}\frac{(z_{ij}-A_{i,j}x_j)^2}{2A_{i,j}}, \\ \text{subject to } \sum_j z_{ij} = b_i, \qquad i = 1,\cdots M, \end{cases} \tag{18}$$

and the iteration is as follows:

$$\begin{aligned} z^{k+1} &= \underset{z}{\text{argmin}}\ E(x^k,z), \qquad \text{subject to } \textstyle\sum_j z_{ij} = b_i. \\ x^{k+1} &= \underset{x}{\text{argmin}}\ E(x,z^{k+1}). \end{aligned}$$

First, by fixing $x = x^k$, we can update z as follows:

$$z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{A_{i,j}}{W_{i,i}}\left(b_i - (Ax^k)_i\right).$$

Then, we can fix $z = z^{k+1}$ and update x. Since

$$\sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} = \frac{1}{2}\sum_j V_{j,j}(x_j - \frac{\sum_i z_{ij}}{V_{j,j}})^2 + \text{constant},$$

x can be updated in the following way:

$$x_j^{k+1} = \frac{\sum_i z_{ij}^{k+1}}{V_{j,j}} = x_j^k + \frac{1}{V_{j,j}}\sum_i \frac{A_{i,j}(b_i - (Ax^k)_i)}{W_{i,i}},$$

which is SART with $w = 1$.

The equivalence of these two optimization problems (15) and (18) is shown below. Since problem (18) is convex, and we can find the minimizer with respect to z for fixed x first and obtain a problem of x:

$$\underset{x}{\text{minimize}} \sum_i \frac{((Ax)_i - b_i)^2}{2W_{i,i}}, \tag{19}$$

which is equivalent to optimization problem (15) SART solves.

6.2 *Relaxation*

In practice, instead of fixing $w = 1$, people use a relaxation of SART [2] to improve the convergence rate. The relaxed version is

$$x_j^{k+1} = x_j^k + \frac{w}{V_{j,j}}\sum_i \frac{A_{i,j}(b_i - (Ax^k)_i)}{W_{i,i}},$$

with a relaxant coefficient w. Inspired by this strategy, we have a relaxation of the EM derivation for SART. We will show that the relaxed EM algorithm is equivalent to solve the unconstrained problem

$$\underset{x,z}{\text{minimize}} E_R(x,z) := \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + \gamma \sum_i \frac{(\sum_j z_{ij} - b_i)^2}{2W_{i,i}}, \tag{20}$$

where $\gamma > 0$, by alternatively minimizing between x and z. First, by fixing $x = x^k$, we can solve the problem of z only, and the analytical solution is

$$z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{\gamma}{1+\gamma}\frac{A_{i,j}}{W_{i,i}}\left(b_i - (Ax^k)_i\right). \quad (21)$$

Then let $z = z^{k+1}$ fixed, and we can find x^{k+1} by solving

$$\underset{x}{\text{minimize}} \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} = \frac{1}{2}\sum_j V_{j,j}(x_j - \frac{\sum_i z_{ij}}{V_{j,j}})^2 + \text{constant},$$

Having z^{k+1} from (21), we can calculate

$$x_j^{k+1} = \frac{\sum_i z_{ij}^{k+1}}{V_{j,j}} = x_j^k + \frac{\gamma}{(1+\gamma)V_{j,j}}\sum_i \frac{A_{i,j}(b_i - (Ax^k)_i)}{W_{i,i}}.$$

Therefore this relaxed EM algorithm is an alternating minimization method. We will show next that the result of this relaxed EM algorithm is the solution to (15).

Because the object function $E_R(x, z)$ in (20) is convex, we can first minimize the function respect to z with x fixed. Then the problem becomes

$$\underset{x}{\text{minimize}}\ E_R(x) = \frac{\gamma}{1+\gamma}\sum_i \frac{((Ax)_i - b_i)^2}{2W_{i,i}}, \quad (22)$$

which is equivalent to (15).

In addition, for finding x^{k+1}, we can choose any x^{k+1} such that the function is decreasing. Then we have

$$x_j^{k+1} = \beta x_j^k + \frac{\beta\gamma}{(1+\gamma)V_{j,j}}\sum_i \frac{A_{i,j}(b_i - (Ax^k)_i)}{W_{i,i}} + (1-\beta)x_j^k \quad (23)$$

where $\beta \in (0, 2)$. This is the classic SART algorithm with parameter $\beta\gamma/(1 + \gamma) \in (0, 2)$ [2].

Remark 1 All the above proofs depend on Lemma 4.1. If some or all elements of the matrix A are negative, the diagonal elements of V and W can be chosen as the summations of the absolute values of elements in columns and rows, which is ℓ_1 norms of columns and rows [7]. Furthermore if we can choose different combinations of V and W to make the lemma valid, the analysis remains the same. Thus, we can choose other different matrices V and W to obtain new algorithms. Actually, there are several methods with different combinations of V and W. Two examples are Cimmino's algorithm and component averaging. Their convergence proofs with changing w can be found in [15].

Cimmino's algorithm [9]: We can choose $V = I$ and W to be the diagonal matrix with diagonal elements $M\|A_{i,\cdot}\|^2$, where $\|A_{i,\cdot}\|$ is the l_2-norm of the i^{th} row.

Component averaging (CAV) [7, 8]: CAV is a novel method based on diagonal weighting and utilizing the sparsity of the matrix A. Denote s_j to be the number of nonzero elements in the j^{th} column of matrix A. Again we can choose $V = I$, and the matrix W is the diagonal matrix with diagonal elements $\sum_{j=1}^{N} s_j A_{i,j}^2$. If all elements of A are nonzero, CAV is exactly the same as Cimmino's algorithm.

7 Noisy Case

In previous sections, we considered the system of linear equations without noise, which is the ideal case. However, noise is unavoidable in applications. In this section, we will consider the noisy problem. The problem is to find x such that

$$Ax + n = b$$

with n being the unknown noise.

Again, we assume that the true solution is x^*, and the result obtained by SART is $\overline{x}$. If b is in the range of A, we have $A\overline{x} = b$, and the following lower bound holds

$$\|\overline{x} - x^*\|_V \geq \|A\overline{x} - Ax^*\|_{W^{-1}} = \|n\|_{W^{-1}}.$$

It is easy to check that Theorem 4.2 is still true for the noisy case, and the residual will decrease in W^{-1}-norm until x^k remains unchanged. From the SART iteration $x^{k+1} = x^k - wV^{-1}A^TW^{-1}(Ax^k - b)$, in order to obtain $x^{k+1} = x^k$, we have $A^TW^{-1}(Ax^k - b) = \mathbf{0}$.

In order to find the upper bound, we first consider a special case where A has full column rank. Then x^k will converge to $\overline{x} = (A^TW^{-1}A)^{-1}A^TW^{-1}b$. Denoting $\|(W^{-1/2}AV^{-1/2})^{-1}\| = \inf\{m : m\|W^{-1/2}AV^{-1/2}x\| \geq \|x\|\}$, we can obtain

$$\begin{aligned}\|\overline{x}-x^*\|_V &= \|V^{1/2}(\overline{x}-x^*)\| \le \|(W^{-1/2}AV^{-1/2})^{-1}\|\cdot\|W^{-1/2}A(\overline{x}-x^*)\| \\ &= \|(W^{-1/2}AV^{-1/2})^{-1}\|\cdot\|A\overline{x}-Ax^*\|_{W^{-1}},\end{aligned}$$

and have to estimate $\|A\overline{x}-Ax^*\|_{W^{-1}}$. $A^TW^{-1}(A\overline{x}-b)=\mathbf{0}$ implies $\langle A\overline{x}-Ax^*, A\overline{x}-b\rangle_{W^{-1}}=0$, and

$$\begin{aligned}\|n\|^2_{W^{-1}} &= \|b-Ax^*\|^2_{W^{-1}} = \|b-A\overline{x}+A\overline{x}-Ax^*\|^2_{W^{-1}} \\ &= \|A\overline{x}-b\|^2_{W^{-1}} + \|A\overline{x}-Ax^*\|^2_{W^{-1}}.\end{aligned}$$

Therefore,

$$\|A\overline{x}-Ax^*\|_{W^{-1}} = \sqrt{\|n\|^2_{W^{-1}} - \|A\overline{x}-b\|^2_{W^{-1}}} \le \|n\|_{W^{-1}}.$$

THEOREM 7.1 *If A has full column rank, we have the following estimation*

$$\|\overline{x}-x^*\|_V \le \|(W^{-1/2}AV^{-1/2})^{-1}\|\cdot\|n\|_{W^{-1}}.$$

Thus, combining (11), we have the estimation for error at each iteration,

$$\begin{aligned}\|x^k-x^*\|_V &\le \|x^k-\overline{x}\|_V + \|\overline{x}-x^*\|_V \\ &\le \frac{\lambda^k}{1-\lambda}\|x^1-x^0\|_V + \|(W^{-1/2}AV^{-1/2})^{-1}\|\cdot\|n\|_{W^{-1}}.\end{aligned}$$

However, when A does not have full column rank, the solution will again depend on the initial guess x^0. From $A^TW^{-1}(Ax-b)=\mathbf{0}$, we have $A^TW^{-1}Ax=A^TW^{-1}b$. If $\tilde{x}$ is the solution with initial guess $x^0=\mathbf{0}$, all the solutions are $\{\tilde{x}+\breve{x}\}$ with $A\breve{x}=\mathbf{0}$. Furthermore $\tilde{x}$ and $\breve{v}$ are orthogonal with respect to V-norm because when $x^0=\mathbf{0}$, all the iterations x^k and $\tilde{x}$ will remain in the range of $V^{-1}A^T$. For other initial guess $x^0\ne\mathbf{0}$ having the decomposition $x^0=\breve{x}+x'$, where $A\breve{x}=\mathbf{0}$ and x' is in the range of $V^{-1}A^T$. It is easy to check that the result $\overline{x}$ will be $\tilde{x}+\breve{x}$.

8 Numerical Experiments

In this section, we provide several numerical experiments to illustrate the convergence of SART, Cimmino's algorithm and CAV for both noise free and noisy cases. All numerical experiments are based on fan-beam computed tomography image reconstruction, and the matrix A representing discrete Radon transform is constructed by Siddon's algorithm [27]. The problem is to reconstruct the image x from the measurements b, which is equivalent with solving $Ax = b$ for noise free case and $Ax + n = b$ for noisy case.

For the first experiment, we consider the noise free case to show the convergence of SART, Cimmino's algorithm and CAV, and compare this with the analysis provided above. To ensure the uniqueness of the solution, a small 32x32 phantom image is chosen with totally 60 views (projections), which is more than enough to make the matrix A having full column rank. For each view, there are 61 measurements.

First, let $w = 1$ fixed, and we solve $Ax = b$ using these three methods for 1000 iterations with the initial guess $x^0 = \mathbf{0}$. From the convergence analysis, we know that the residual $b - Ax^k$, difference between two iterations $x^k - x^{k-1}$, and error $x^* - x^k$ are exponential decreasing in W^{-1}-norm , V-norm and V-norm respectively. The numerical results are shown in Figure 1.

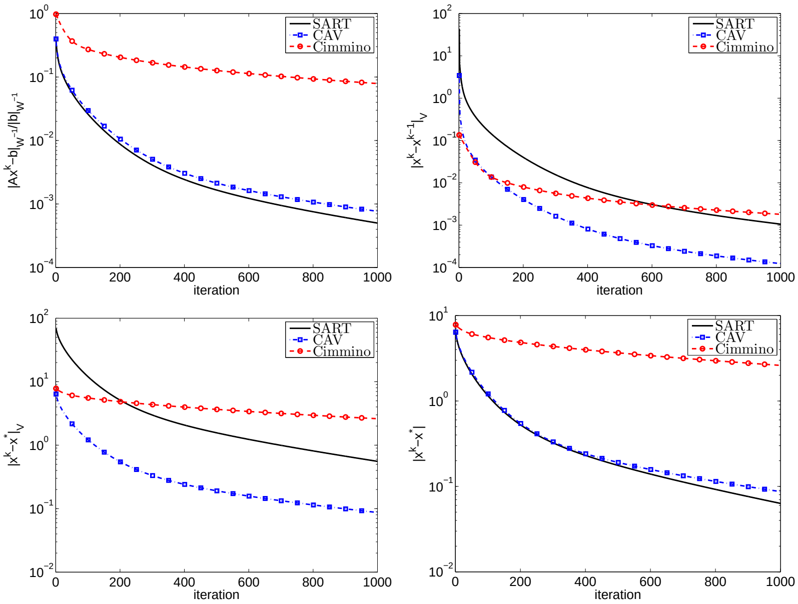

Figure 1. Decays of residual in W^{-1}-norm, difference between two iterations in V-norm, error in V-norm and l_2-norm.

From Figure 1, we can easily see the exponential decay of residual, difference between two iterations and error in ellipsoidal norms for these three methods. Since for these three methods, the ellipsoidal norms are different, we also show the error in l_2-norm for these methods at each iteration in Figure 1.

The comparing results in Figure 1 show that SART and CAV converge faster than Cimmino's algorithm for this special $w = 1$. Then we perform the same numerical experiment with different w's for different methods. Though we only show the convergence for $w \in (0, 2)$ in section 4, $w = 2,\ 2.35,\ 50$ are chosen for SART, CAV and Cimmino's algorithm respectively. The decays of residual, difference between two iterations and error in corresponding ellipsoidal norms are shown in Figure 2.

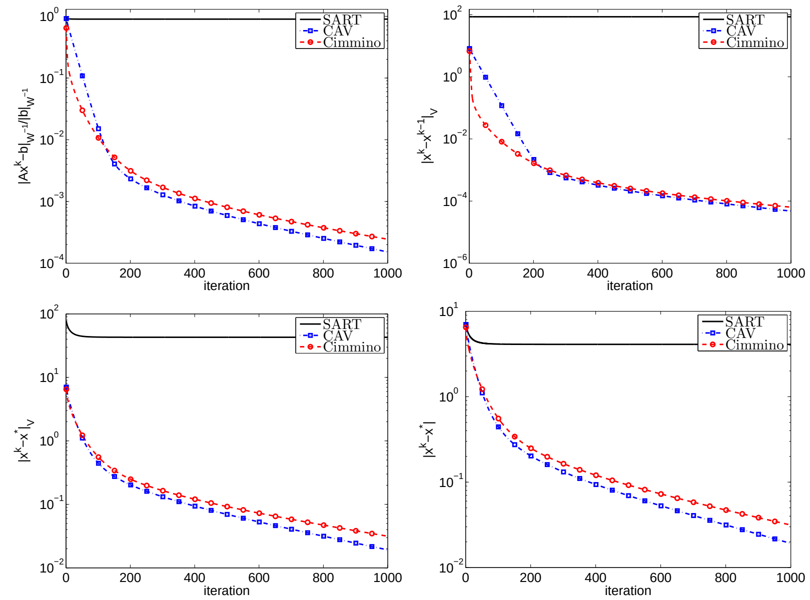

Figure 2. Decays of residual in W^{-1}-norm, difference between two iterations in V-norm, error in V-norm and l_2-norm.

These figures show that Cimmino's algorithm and CAV will converge even when $w > 2$, while for SART, the convergence only occurs when $w \in (0, 2)$. In addition, we have the reconstruction results for these methods in Figure 3.

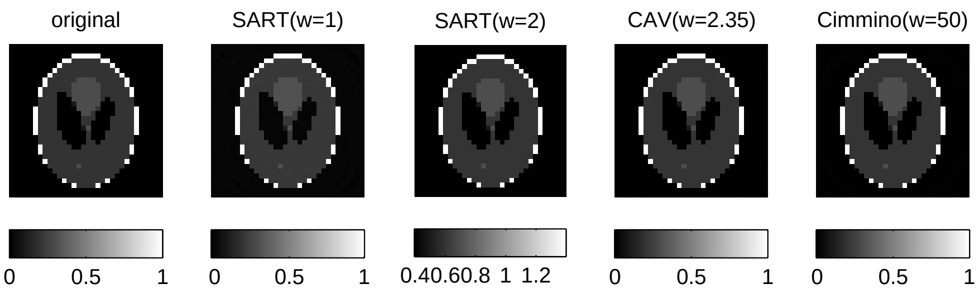

Figure 3. Reconstruction results of a 32x32 image with $w = 1,\ 2,\ 2.35,\ 50$ for SART, CAV, Cimmino's algorithm respectively.

Since the solution is unique, x^k will converge to the original true image if the algorithm converges (SART with $w = 1$, CAV with $w = 2.35$, and Cimmino's algorithm with $w = 50$). For SART with $w = 2$, it does not converge, but the result image looks similar to the original image, without considering the color bar at the bottom. In fact, the result image of SART with $w = 2$ is shifted by a constant c. From the remark after Lemma 4.1, we know that $\mathbf{1}$ is in the null space of $V - A^TW^{-1}A$. Thus, if $x^k \approx x^* + c\mathbf{1}$, we will have

$$x^{k+1} = x^k - 2V^{-1}A^TW^{-1}(Ax^k - b) \approx x^* + c\mathbf{1} - 2V^{-1}A^TW^{-1}Ac\mathbf{1} = x^* - c\mathbf{1}.$$

Similarly $x^{k+2} \approx x^{k+1} + 2c\mathbf{1} \approx x^k$ and this explains why the residual, difference between two iterations and error do not decay in corresponding norms, and the image reconstructed using SART is shifted by a constant.

However, CAV and Cimmino's algorithm will still converge even when $w > 2$, because we have the following more accurate constraint for w with any pair of V and W, $0 < w < 2/\rho(V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}})$, where $\rho(G)$ is the largest eigenvalue of the matrix G. For SART, $\rho(V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}) = 1$ because $V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}} \prec I$ and $V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}V^{\frac{1}{2}}\mathbf{1} = V^{\frac{1}{2}}\mathbf{1}$, and we have the constraint $0 < w < 2$. While for CAV and Cimmino's algorithm, the upper-bounds depend on the matrix A. For this special A in the numerical experiment, $\rho(V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}) < 1$ for CAV and Cimmino's algorithm. Thus from the figures, we can see that the residual, difference between two iterations and error still decrease exponentially in corresponding ellipsoidal norms. For SART, if w is not fixed for all steps, w can be greater than 2 for some cases, for convergence of SART with different relaxation coefficients in different steps, please see [24].

If the measurements are not sufficient to ensure that the solution to the system is unique, the result will depend on the initial guess. If we choose only 15 views out of the 60 views, then the number of measurements is only 915 while the number of pixels is 1024. This is a underdetermined system, having infinite many solutions. In the following Figure 4 we can find the result with three different initial guesses.

For a larger 256x256 image, if we choose 180 views with 301 measurements in each view, it still leads to an underdetermined system. We choose three different initial guesses and repeat the experiment. The reconstructed results after 1000 iterations are in Figure 5.

These two figures show the importance of initial guess when the uniqueness of the system is not satisfied. We can not reconstruct the original image without any prior knowledge about it even in the noise free case, because there are many solutions for the problem $Ax = b$ and SART only provides us with one solution, which depends on the initial guess. Usually the initial guess for

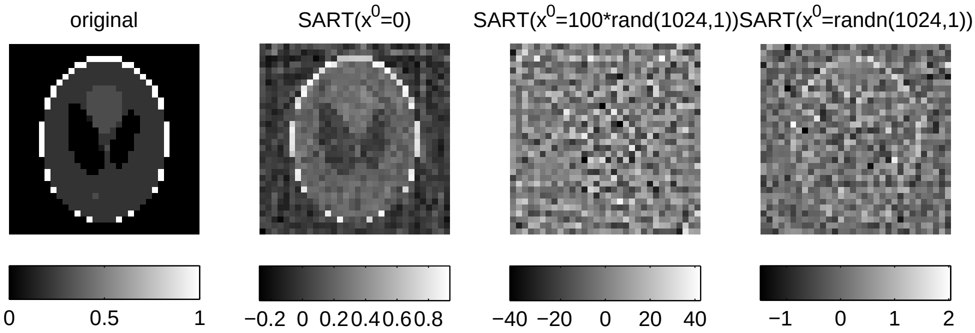

Figure 4. Reconstruction results of a 32x32 image with different initial guesses.

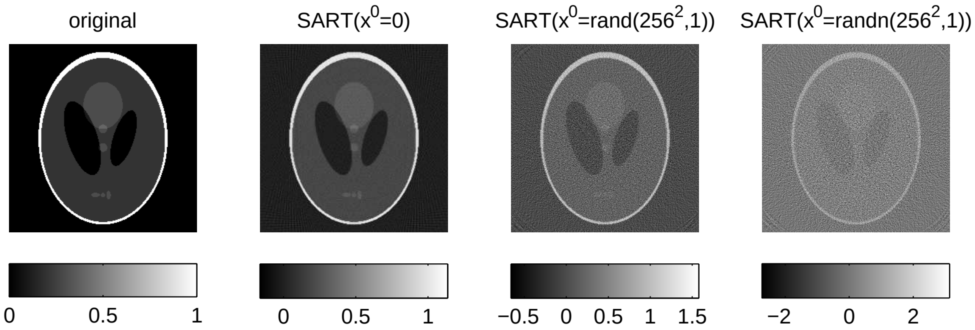

Figure 5. Reconstruction results of a 256x256 image with different initial guesses.

CT reconstruction is **0** or the result from FBP, which are not the best choices for initial guess. Finding the best initial guess is as difficult as finding the original image. Thus for the insufficient measurements case, we have to use some regularization methods such as total variation (TV) minimization and compressed sensing. Some methods combining SART and compressed sensing (CS) or TV minimization [23, 28, 30, 33] are proposed for CT reconstruction to reduce the number of measurements furthermore.

For the last experiment, we consider the noisy case and the system to be solved is

$$Ax + n = b.$$

Here, we choose the small 32x32 image again. Assume that n is the additive Gaussian white noise with zero means and different variances. We choose the variances increasing equally from 0.01 to 1 and show the V-norm of the error, comparing with the bound given in the last section. As is shown in Figure 6, the error is quite close to the bound.

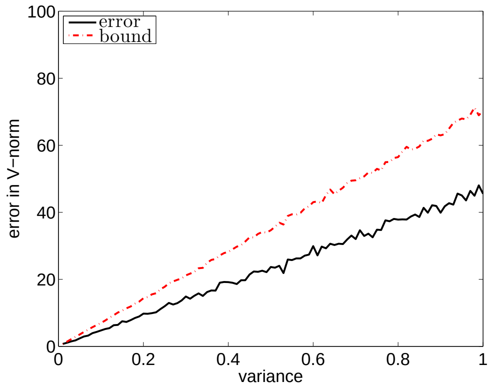

Figure 6. Error of the results with different levels of noise.

9 Conclusion

In this paper, we proposed several novel derivations for SART from optimization and statistical points of view and extended the analysis to two other Landweber-like schemes: I) SART can be derived from linearized Bregman iteration for primal problem (2) and gradient descent with unit step for dual problem (14); II) The connection between SART and EM is provided by showing that SART is an EM algorithm and alternating minimization method. Furthermore, the exponential decays of residual and error in corresponding ellipsoidal norms are given. For the noisy case, we provide an error bound for the result.

Acknowledgement

I would like to thank Professor Luminita A. Vese for valuable discussions and suggestions. This work is supported by the Center for Domain-Specific Computing (CDSC) funded by the NSF Expedition in Computing Award CCF-0926127.

References

[1] A. Andersen, *Algebraic reconstruction in CT from limited views*, IEEE Transactions on Medical Imaging 8 (1989), pp. 50–55.
[2] A. Andersen and A. Kak, *Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm*, Ultrasonic Imaging 6 (1984), pp. 81–94.
[3] C. Atkinson and J. Soria, *An efficient simultaneous reconstruction technique for tomographic particle image velocimetry*, Experiments in Fluids 47 (2009), pp. 553–568.
[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004).
[5] L. Bregman, *The relaxation method for finding common points of convex sets and its application*

to the solution of problems in convex programming, USSR Computational Mathematics and Mathematical Physics 7 (1967), pp. 200–217.

[6] J. Cai, S. Osher, and Z. Shen, *Linearized Bregman iterations for compressed sensing*, Mathematics of Computation 78 (2009), pp. 1515–1536.

[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 24 (2002), pp. 40–58.

[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), pp. 777–808.

[9] G. Cimmino, *Calcolo approssimato per le soluzioni dei sistemi diequazioni lineari*, La Ricerca Scientifica, II 9 (1938), pp. 326–333.

[10] A. Dempster, N. Laird, and D. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society Series B 39 (1977), pp. 1–38.

[11] R. Gordon, R. Bender, and G. Herman, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography*, Journal of Theoretical Biology 29 (1970), pp. 471–481.

[12] P.C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1998).

[13] G. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction From Projection*, Springer (2009).

[14] M. Jiang and G. Wang, *Convergence of the simultaneous algebraic reconstruction technique (SART)*, IEEE Transaction on Image Processing 12 (2003), pp. 957–961.

[15] ———, *Convergence studies on iterative algorithms for image reconstruction*, IEEE Transactions on Medical Imaging 22 (2003), pp. 569–579.

[16] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Society of Industrial and Applied Mathematics (2001).

[17] A. Kirsch, *An introduction to the mathematical theory of inverse problems*, Applied Mathematical Sciences, Springer (1991).

[18] L. Landweber, *An iteration formula for fredholm integral equations of the first kind*, The American Journal of Mathematics 73 (1951), pp. 615–624.

[19] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer (1999).

[20] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling and Simulation 4 (2005), pp. 460–489.

[21] S. Osher, Y. Mao, B. Dong, and W. Yin, *Fast linearized Bregman iteration for compressed sensing and sparse denoising*, Communications in Mathematical Sciences 8 (2010), pp. 93–111.

[22] X. Pan, E.Y. Sidky, and M. Vannier, *Why do commercial ct scanners still employ traditional, filtered back-projection for image reconstruction?*, Inverse Problems 25 (2009), p. 123009.

[23] Y. Pan, R. Whitaker, A. Cheryauka, and D. Ferguson, *Feasibility of GPU-assisted iterative image reconstruction for mobile C-arm CT*, Proceedings of International Society for Photonics and Optonics (SPIE) 7258 (2009), p. 72585J.

[24] G. Qu, C. Wang, and M. Jiang, *Necessary and sufficient convergence conditions for algebraic image reconstruction algorithms*, IEEE Transactions on Image Processing 18 (2009), pp. 435–440.

[25] L. Shepp and B. Logan, *The Fourier reconstruction of a head section*, IEEE Transaction on Nuclear Science 21 (1974), pp. 21–34.

[26] L. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Transaction on Medical Imaging 1 (1982), pp. 113–122.

[27] R. Siddon, *Fast calculation of the exact radiological path for a three-dimensional CT array*, Medical Physics 12 (1986), pp. 252–255.

[28] J. Tang, B. Nett, and G. Chen, *Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms*, Physics in Medicine and Biology 54 (2009), pp. 5781–5804.

[29] J. Wang and Y. Zheng, *On the convergence of generalized simultaneous iterative reconstruction algorithms*, IEEE Transaction on Image Processing 16 (2007), pp. 1–6.

[30] M. Yan and L.A. Vese, *Expectation maximization and total variation based model for computed tomography reconstruction from undersampled data*, Proceedings of International Society for Photonics and Optonics (SPIE) 7961 (2011), p. 79612X.

[31] W. Yin, *Analysis and generalizations of the linearized bregman method*, SIAM J. on Imaging Sciences 3 (2010), pp. 856–877.

[32] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for l1- minimization with applications to compressed sensing*, SIAM J. on Imaging Sciences 1 (2008), pp. 143–168.
[33] H. Yu and G. Wang, *SART-type image reconstruction from a limited number of projections with the sparsity constraint*, International Journal of Biomedical Imaging (2010), article ID 934847.