Convergence Analysis of SART by Bregman Iteration and Dual Gradient Descent

Ming Yan*

December 10, 2010

Abstract

Simultaneous algebraic reconstruction technique (SART) [1, 2] is an iterative method for solving inverse problems of the form $Ax = b$. This type of problems arises for example in computed tomography reconstruction, when A is obtained from the discrete Radon transform. In this paper, we provide two new methods for the derivation of the SART method. Using these, we also prove in two different ways the convergence of the simultaneous algebraic reconstruction technique. The first approach uses the linearized Bregman iteration, while the second approach uses the dual gradient descent method. These novel proofs 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 estimate is given. Several numerical experiments for computed tomography are provided to demonstrate the convergence results in practice.

Keywords: simultaneous algebraic reconstruction technique, Bregman iteration, dual gradient descent, image reconstruction, component averaging, Cimmino's algorithm

1 Introduction

As a group of methods for reconstructing two dimensional and three dimensional images from the projections of the object, iterative reconstruction has many applications such as in computerized tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI). This technique is quite different from the filtered back projection (FBP) method [20], which is the most commonly used algorithm in practice by manufacturers. The main advantages of the iterative reconstruction technique over FBP are insensitivity to noise and flexibility [13]. 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.

There are many available algorithms for iterative reconstruction of the solution of an inverse problem. The inverse problem to be solved is based on the system of linear equations

$$Ax = b,$$

where $x = (x_1, \cdots, x_N)^T \in \mathbf{R}^N$ is the unknown (for example, the image to be reconstructed from projections expressed as a long vector), b is the given measurement with $b = (b_1, \cdots, b_M)^T \in \mathbf{R}^M$, and A is a $M \times N$ matrix describing the transformation from x to the measurements. This matrix A is different for different purposes. For example, in computed tomography, A represents the discrete Radon transform, with each row describing an integral along one straight line, and having all the elements nonnegative.

Some examples of iterative reconstruction algorithms are expectation maximization (EM) [21], algebraic reconstruction techniques (ART) [9, 10], and component averaging methods (CAV) [6]. Simultaneous algebraic reconstruction technique (SART) [1, 2], as a refinement of ART, is also widely

*Department of Mathematics, University of California, Los Angeles, CA 90095. Email: yanm@math.ucla.edu

used [3, 18, 27]. Furthermore, the convergence analysis of SART and CAV are studied by Jiang and Wang et. al. [11, 12, 19], Wang and Zheng [24], Censor and Elfving [6].

The Bregman iteration [5] is a technique used to solve a constrained optimization problem by solving a sequence of unconstrained problems [5]. Osher et al. used this in image restoration [16] and the technique is now widely used in compressed sensing and image processing to solve $l1$-minimization problems. In addition, a linearized version of Bregman iteration, named linearized Bregman iteration is proposed by Osher et al. [17] and Yin et al. [26].

We will provide in this paper a novel convergence analysis of the SART method from the linearized Bregman iteration of an optimization problem and from the gradient descent method with unit step applied to the dual problem. The exponential decay of the residual and of the difference with the exact solution is shown in corresponding norms. Furthermore, for the noisy case, a bound of the error is provided in corresponding norm.

The organization of the paper is as follows: in section 2, we will give an introduction to the SART algorithm. The concept of Bregman iteration and linearized Bregman iteration, along with the connection of SART and linearized Bregman iteration of an optimization problem, will be provided in section 3. Then we will show the convergence analysis of the SART method using linearized Bregman iteration and dual gradient descent in sections 4 and 5 respectively. Also, we show that the convergence analysis can be applied to a general Landweber scheme [14]. Furthermore, we discuss the noisy case in section 6. In order to illustrate the convergence results shown in this paper, we present numerical experiments for both noise free and noisy cases in section 7, corresponding to the computed tomography case. Finally, we end up with a conclusion and remark section.

2 Simultaneous Algebraic Reconstruction Technique

In this section, we provide an introduction to the simultaneous algebraic reconstruction technique (SART) [1, 2]. Recall from section 1, we have the system of linear equations

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

Here, we have the assumption that all elements in A are nonnegative (this is a property that must hold when A is obtained from the discrete Radon transform in computed tomography). Even if this assumption is not satisfied, we still obtain the same convergence analysis without any difficulty (for more details, see the remark in the last section).

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, and the existence of the solution is guaranteed. 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

$$A_{i,+} = \sum_{j=1}^{N} A_{i,j} \qquad \text{for } i = 1, \cdots, M,$$
$$A_{+,j} = \sum_{i=1}^{M} A_{i,j} \qquad \text{for } j = 1, \cdots, N.$$

$A_{i,+}$ is the summation of all elements in the i^{th} row, and $A_{+,j}$ is the summation of all elements in the j^{th} column. In addition, there is an assumption on $A_{i,+}$ and $A_{+,j}$ that $A_{i,+} > 0$ and $A_{+,j} > 0$. Actually if $A_{i,+} = 0$, the i^{th} measurement is 0 for any x. $A_{+,j} = 0$ means that change in the j^{th} component of x cannot be detected in the measurements.

Let V be the diagonal matrix with diagonal elements $A_{+,j}$, and W be the diagonal matrix with diagonal elements $A_{i,+}$. The SART method proposed in [1, 2] is

$$x^{k+1} = x^k + wV^{-1}A^TW^{-1}(b - Ax^k), \tag{2}$$

for $k = 0, 1, \cdots$, where w is a relaxation parameter in $(0, 2)$, and the starting point is x^0.

3 SART from Bregman Iteration

Before deriving SART from the Bregman iteration, we provide some definitions for later use. First we recall the definition of ellipsoidal norm for vectors. Let G be an $n \times n$ symmetric positive definite matrix. The ellipsoidal norm of $x \in R^n$ is defined as follows:

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

Since we may obtain different ellipsoidal norms for different matrices, we indicate the positive definite matrix in the notation of these norms. The ellipsoidal norm with matrix G is named as G-norm. We will use two ellipsoidal norms $\|x\|_V$ and $\|y\|_{W^{-1}}$, named V-norm and W^{-1}-norm respectively.

From the convergence analysis of SART in [12], SART converges to the solution of $Ax = b$ with the least V-norm if the initial guess $x^0 = 0$. We consider the following problem which depends on the initial guess x^0,

$$\begin{cases} \underset{x}{\text{minimize}} & \|x - x^0\|_V^2, \\ \text{subject to} & Ax = b. \end{cases} \tag{3}$$

This is a convex constrained optimization problem and the existence and the uniqueness of a solution is guaranteed. 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,

$$\underset{x}{\text{minimize}}\, \mu\|x - x^0\|_V^2 + \frac{1}{2}\|Ax - b\|_{W^{-1}}^2, \tag{4}$$

where μ is a positive parameter.

The solution of (4) converges to the solution of (3) as $\mu \to 0$ [15]. Thus, we can find an approximate solution of (3) by solving (4) with a sufficient small μ. However, small μ will slow down many algorithms.

Recently, the Bregman iteration method for solving constrained problems by solving a sequence of unconstrained problems is proposed by Osher et al. [16] in image processing. Actually this idea dates back to 1967, provided by Bregman [5].

Instead of solving one unconstrained problem to obtain an approximate solution of problem (3), we solve several unconstrained problems iteratively, and the result also converges to the solution of the unconstrained problem.

We define $J(x) = \mu\|x - x^0\|_V^2$ and the Bregman distance by

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

with $p \in \partial J(x^2)$, a subgradient of J at x^2. In general, the Bregman distance is not a distance in the common sense. However, for this special J in this paper, the Bregman distance is a distance in the common sense, and we have $D_J^p(x^1, x^2) = \mu\|x^1 - x^2\|_V^2$.

The Bregman iteration is as follows

$$\begin{aligned} x^{k+1} &= \underset{x}{\text{argmin}}\ D_J^p(x, x^k) + \frac{1}{2}\|Ax - b\|_{W^{-1}}^2 \\ &= \underset{x}{\text{argmin}}\ \mu\|x - x^k\|_V^2 + \frac{1}{2}\|Ax - b\|_{W^{-1}}^2, \end{aligned} \tag{5}$$

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 (5), it follows that

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

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

$$(2\mu V + A^T W^{-1} A)x^{k+1} = A^T W^{-1} b + 2\mu V x^k. \tag{6}$$

The matrix in the left-hand-side remains unchanged while the right-hand-side changes for each iteration.

Furthermore, we can linearize the second quadratic term in (5) and obtain the linearized Bregman iteration as follows,

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\ \mu\|x - x^k\|_V^2 + \langle Ax^k - b, Ax\rangle_{W^{-1}} + \frac{1}{2\alpha}\|x - x^k\|_V^2. \tag{7}$$

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 $w = \frac{1}{2\mu+\frac{1}{\alpha}}$, we thus observe that the SART method 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 Linearized Bregman Iteration

In this section, we will show the convergence of SART from linearized Bregman iteration. First we state the following important lemma [11], which will be used several times.

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

Remark: In addition, we have $\|Ax\|_{W^{-1}} \le \|x\|_V$. Furthermore, $V - A^T W^{-1} A$ and $W - AV^{-1}A^T$ are not positive definite, because x with same constant value for each component is in the null spaces of $V - A^T W^{-1} A$ and $W - AV^{-1}A^T$.

Firstly, we show that the residual is decreasing with respect to the W^{-1}-norm by the following theorem, and we give a different proof from [12].

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

$$\|Ax^{k+1} - b\|_{W^{-1}}^2 + (\frac{2}{w} - 1)\|x^{k+1} - x^k\|_V^2 \le \|Ax^k - b\|_{W^{-1}}^2, \tag{8}$$

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

Proof: The linearized iteration (7) is equivalent to

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

and from the optimality of x^{k+1}, we have

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

which is one step of SART.

It is easy to verify that

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

which is equivalent to

$$\frac{1}{w}\|x^{k+1}-x^k\|_V^2+\langle Ax^k-b,Ax^{k+1}-Ax^k\rangle_{W^{-1}}=0. \tag{9}$$

In addition, we have

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

We plug this into equation (9), and we have

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

From the remark after Lemma 4.1, we have $\|Ax^{k+1}-Ax^k\|_{W^{-1}}^2\le\|x^{k+1}-x^k\|_V^2$.

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 constraint problem (3). We show the convergence of x^k first. Since

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

denote $r^k=x^{k+1}-x^k$. It follows that

$$\begin{aligned}V^{\frac{1}{2}}r^k&=V^{\frac{1}{2}}(I-wV^{-1}A^TW^{-1}A)V^{-\frac{1}{2}}V^{\frac{1}{2}}r^{k-1}\\ &=(I-wV^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}})V^{\frac{1}{2}}r^{k-1}.\end{aligned} \tag{10}$$

In order to show the exponential decay of $\|r^k\|_V$ or $\|V^{\frac{1}{2}}r^k\|$, we have to show that

$$0\prec wV^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}\prec 2I \tag{11}$$

during the iterations. From Lemma 4.1 we have $A^TW^{-1}A\preceq V$, which is equivalent to

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

thus for $w\in(0,2)$,

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

However, $V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}\succ 0$ is not always true. If A has nontrivial null space $\mathcal{N}(A)$, we can choose 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 subset of R^N, the image of $V^{-\frac{1}{2}}A^T$, we have $V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}} \succ 0$.

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 $V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}} \succ 0$ is satisfied during the iterations.

Therefore, for $w \in (0, 2)$, from the exponential decay of $\|r^k\|_V$ in (10) and (11), there exists a constant $\lambda \in (0, 1)$ such that $\|x^{k+1} - x^k\|_V \le \lambda\|x^k - x^{k-1}\|_V \le \lambda^k\|x^1 - x^0\|_V$. Therefore

$$\|x^k - \overline{x}\|_V \le \sum_{i=k}^{\infty} \|x^{i+1} - x^i\|_V \le \sum_{i=k}^{\infty} \lambda^i\|x^1 - x^0\|_V = \frac{\lambda^k}{1-\lambda}\|x^1 - x^0\|_V. \tag{12}$$

Theorem 4.3 *Assume that x^* is the solution of $Ax = b$ with the least $\|x - x^0\|_V$, and $x^k \to \overline{x}$. Then we have the following estimate:*

$$\mu\|\overline{x} - x^0\|_V^2 \le \mu\|x^* - x^0\|_V^2 + \frac{1}{\alpha}\langle \overline{x} - x^0, x^* - \overline{x}\rangle_V. \tag{13}$$

Proof: From the updating of x^k, we have

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

Since $w = \frac{1}{2\mu + \frac{1}{\alpha}}$, we have

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

From the positivity of Bregman distance, we have

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

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.

From the updating of x^k, we have

$$x^{k+1} - x^k = wV^{-1}A^TW^{-1}(b - Ax^k). \tag{14}$$

Multiply by A and we have

$$Ax^{k+1} - b = Ax^k - b - wAV^{-1}A^TW^{-1}(Ax^k - b) = (I - wAV^{-1}A^TW^{-1})(Ax^k - b),$$

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

Theorem 4.4 *x^k converges to x^*, or $\overline{x} = x^*$.*

Proof: From the proof of the previous theorem, the 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 $A(\overline{x} - x^*) = 0$. Furthermore, $\overline{x} = x^0 + wV^{-1}A^TW^{-1}\sum_{j=0}^{\infty}(b - Ax^j)$, 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 (13) 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 with the least $\|x - x^0\|_V$. ∎

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 of linearized Bregman iteration for compressive sensing and gradient descent with unit step for dual problem is shown in [25] by Yin. Here we can also derive the convergence by considering the gradient descent method of the dual problem. Assume that the primal 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 constraint problem (3).

The Lagrangian function [4] is

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

with y being the Lagrangian coefficient 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

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

Plugging this into the Lagrangian function, we have

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

Therefore, the dual problem is

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

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

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

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

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

The convergence can be derived from gradient descent. Since

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

we have

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

$F(y^k)$ will decreasing until y^k remain 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

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

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

Remark: All the above proofs depend on Lemma 4.1. 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 can be found in [11], different from the present ones.

Cimmino's algorithm [8]: 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) [6, 7]: 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 $\sum_{j=1}^{N} s_jA_{i,j}^2$. If all elements of A are nonzero, CAV is exactly the same as Cimmino's algorithm.

6 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

$$
Ax+e=b
$$

with e being the unknown noise. Assume that it is a Gaussian noise with different variances for different components. Then the probability based on Gaussian probability distribution function is

$$P(b|Ax) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi A_{i,+}}} e^{-\frac{(b_i - a_i x)^2}{2A_{i,+}}},$$

where a_i is the i^{th} row of A. In the case of computed tomography reconstruction (CT), if $A_{i,+}$ (the length of intersection of X-ray with the domain) is large, the variance of the noise will also be large.

Thus, we can consider to find x such that the probability is largest. Instead of maximizing the probability, we minimize

$$-\log P(b|Ax) = \sum_{i=1}^{M} A_{i,+}^{-1}(a_i x - b_i)^2 + constant = \|Ax - b\|_{W^{-1}} + constant.$$

Using steepest descent method, we recover the SART iteration

$$x^{n+1} = x^n - wV^{-1}A^T W^{-1}(Ax^n - b).$$

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

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

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 $x^{k+1} = x^n - wV^{-1}A^T W^{-1}(Ax^n - b)$, in order to obtain $x^{k+1} = x^k$, we have $V^{-1/2}A^T W^{-1}(Ax^k - b) = 0$.

For finding the upper bound, we first consider a special case when A has full column rank. Then x^k will converge to $\overline{x} = (A^T W^{-1} A)^{-1} A^T W^{-1} b$. Denote $\|(W^{-1/2}AV^{-1/2})^{-1}\| = \inf\{M : M\|W^{-1/2}AV^{-1/2}x\| \geq \|x\|\}$. Then, we have

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

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

$$\|e\|_{W^{-1}}^2 = \|b - Ax^*\|_{W^{-1}}^2 = \|b - A\overline{x} + A\overline{x} - Ax^*\|_{W^{-1}}^2 = \|A\overline{x} - b\|_{W^{-1}}^2 + \|A\overline{x} - Ax^*\|_{W^{-1}}^2.$$

Therefore,

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

Theorem 6.1 *If A has full column rank, then we have the following estimate*

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

Therefore, combining (12) we have the estimate for error at each iteration,

$$\|x^k - x^*\|_V \leq \|x^k - \overline{x}\|_V + \|\overline{x} - x^*\|_V \leq \frac{\lambda^k}{1-\lambda}\|x^1 - x^0\|_V + \|(W^{-1/2}AV^{-1/2})^{-1}\| \cdot \|e\|_{W^{-1}}.$$

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

7 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 the discrete Radon transform is constructed by Siddon's algorithm [22]. The problem is to reconstruct the image x from the measurements b, which is equivalent with solving $Ax = b$.

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. We choose a small 32x32 phantom image, and for the measurements, we take views for every 6 degrees, and 60 views totally. For each view, there are 61 measurements, which is enough to make sure that the solution to the system is unique.

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

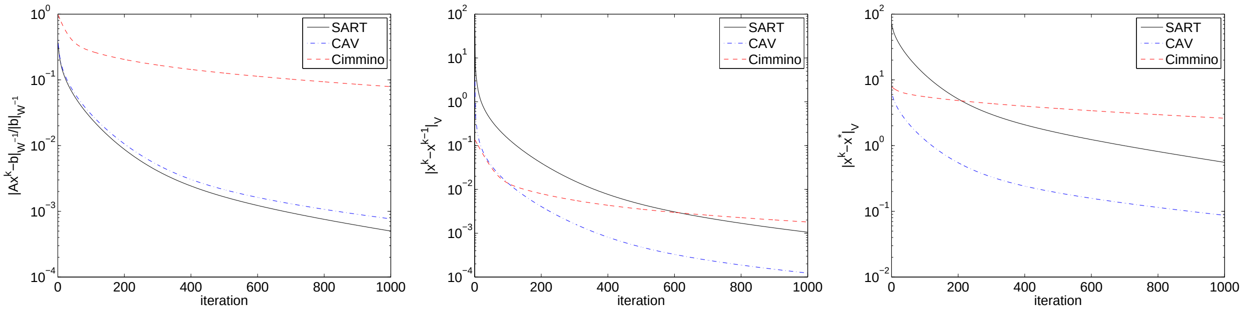

Figure 1: Decay of residual, of the norm difference between two iterations and error in corresponding ellipsoidal norms.

From Figure 1, we can easily see the exponential decay of the residual, of the difference between two iterations and of the 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 the following Figure 2.

Comparing the decays of these three methods, we find that the convergences of SART and CAV are much faster than of Cimmino's algorithm for this special $w = 1$. Then we perform the same numerical experiment with different w for different methods. Though we only show the convergence for $w \in (0, 2)$, we choose $w = 2, 2.35, 50$ for SART, CAV and Cimmino's algorithm respectively. The decay of residual, difference between two iterations and error in corresponding ellipsoidal norms are shown in Figure 3. The error in l_2-norm is in Figures 4.

These figures show that Cimmino's algorithm and CAV will converge even when $w > 2$, while for SART, the convergence works only for $w \in (0, 2)$. In addition, we have the reconstruction results for these three methods in Figure 5.

From these results, we can see that the reconstructions by these three methods are quit close to the original true image if we don't consider the color bar at the bottom. In fact, the range result of SART is moved by a constant ≈ -0.1280. From the remark after Lemma 4.1, we know that x with constant value for each component is in the null space of $V - A^TW^{-1}A$. Thus, if $x^k \approx x^* + c$, with c being a

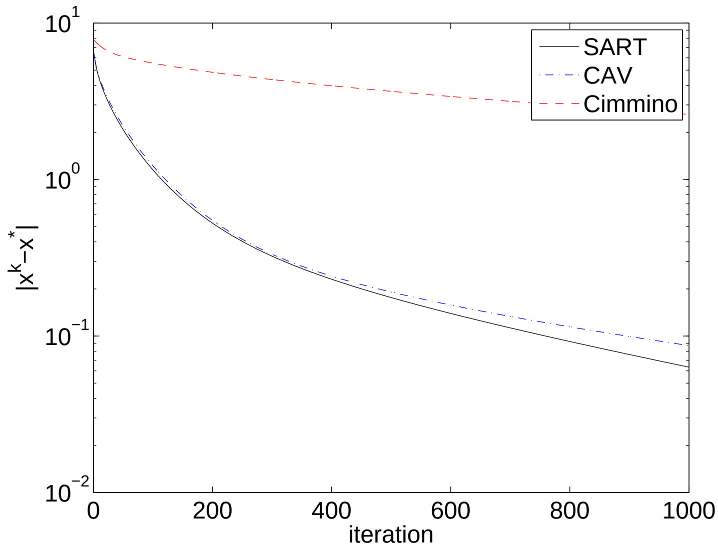

Figure 2: Decay of error in l_2-norm for three methods

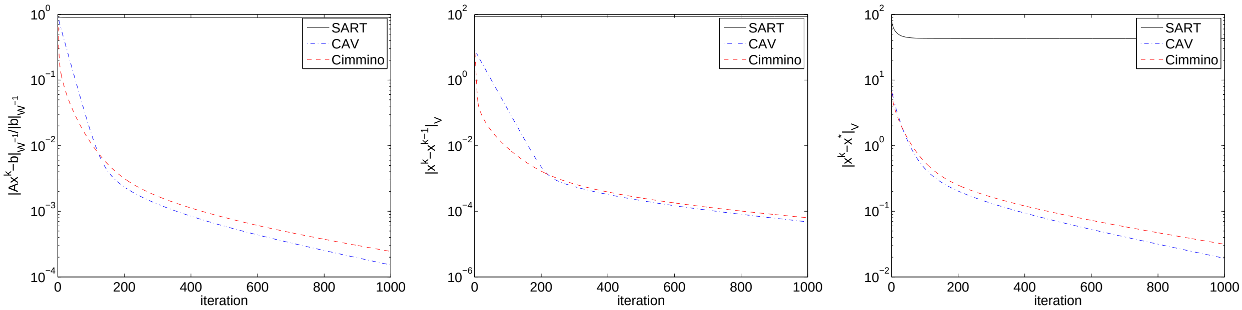

Figure 3: Decay of residual, difference between two iterations and error in ellipsoidal norms

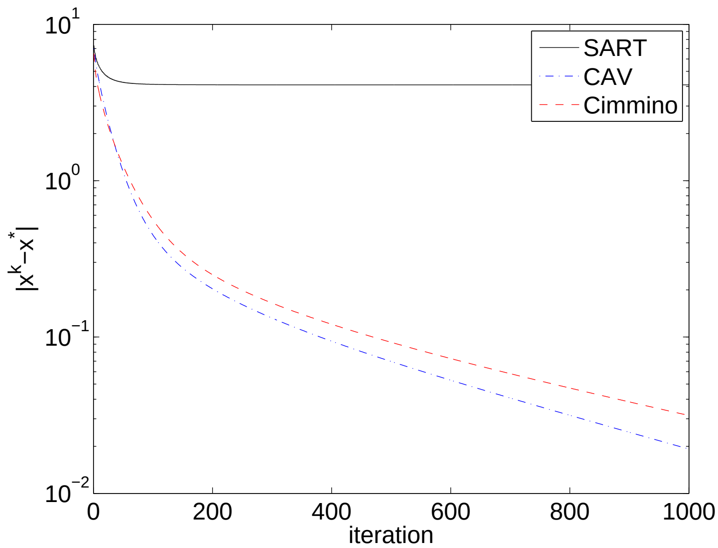

Figure 4: Decay of error in $l2$-norm for three methods

vector with the same value for each component, we will have

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

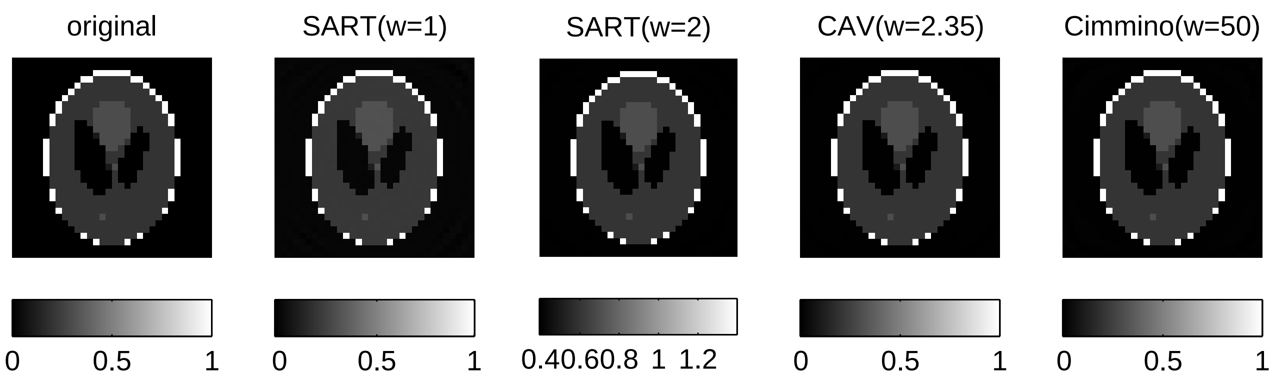

Figure 5: Reconstruction results of a 32x32 image with $w = 1, 2, 2.35, 50$ for SART, CAV, Cimmino's algorithm respectively.

Therefore $x^{k+2} \approx x^{k+1} + 2c \approx x^k$, and this explains why the residual,the difference between two iterations and the 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 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, $\rho(V^{-\frac{1}{2}}A^TW^{-1}AV^{-\frac{1}{2}}) = 1$ and we have the constraint $0 < w < 2$, while for CAV and Cimmino's algorithm, the upperbound depends 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 the residual, difference between two iterations and error still decrease exponentially in corresponding ellipsoidal norm. For convergence of SART with different relaxation coefficents in different steps, see [19].

If the measurements are not sufficient to insure that the solution to the system is unique, the result will depend on the initial guess. If we choose only 15 views, one for 24 degrees, 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 6 we can find the result with three different initial guesses.

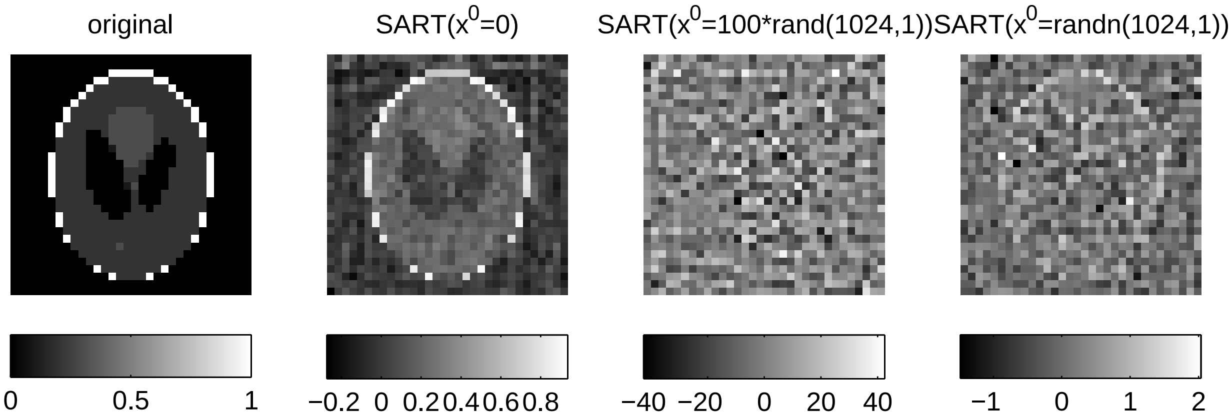

Figure 6: Reconstruction results of a 32x32 image with different initial guesses.

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 we repeat the experiment. The reconstructed results after 1000 iterations are in Figure 7.

If the initial guess is not chosen properly, we will not obtain something useful, as the third one ($x^0 = 100 * \text{rand}(1024,1)$) in Figure 6. In another way, if there is a way to choose a better initial guess,

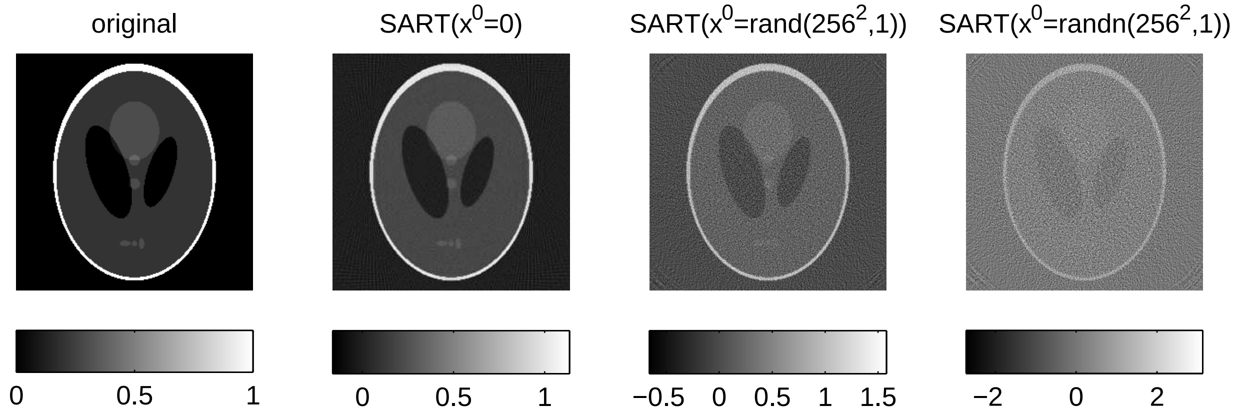

Figure 7: Reconstruction results of a 256x256 image with different initial guesses.

the result will be improved. From these two figures, we can see that $x^0 = 0$ is a better guess comparing with other two guesses. However, $x^0 = 0$ is not the best one and how to choose a good initial guess is still a difficult problem.

The last two experiments with insufficient measurements show that 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 and SART only provides us one solution, which depends on the initial guess. Thus for the insufficient measurements case, we have to use some regularization method such as total variation (TV) minimization and compressed sensing. Some methods combining SART and compressed sensing (CS) or TV [18, 23, 27] are proposed for CT reconstruction to reduce the number of measurements further more.

We will consider the noisy case and the system to be solved is

$$Ax + e = b.$$

Here, we choose the small 32x32 image again. Assume that e is the additive Gaussian 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 8, the error is quite close to the bound.

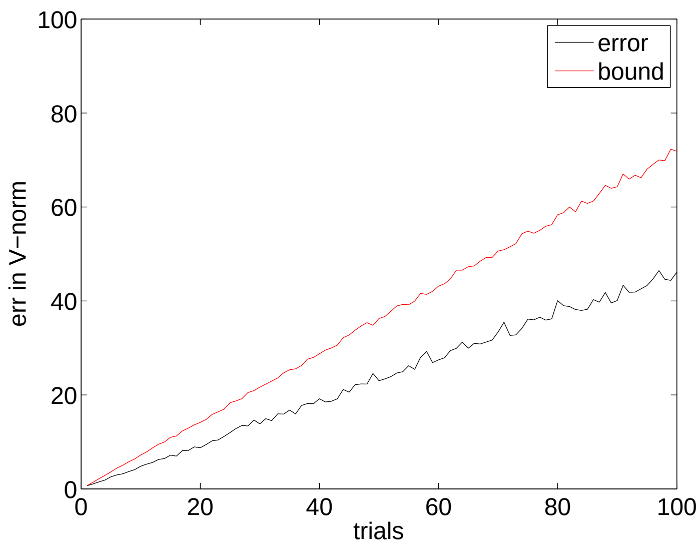

Figure 8: Error of the results with different noise.

8 Conclusion and Remarks

In this paper, we proposed two novel approaches for the convergence of the SART method and extended the analysis to two other Landweber-like schemes. One approach uses the linearized Bregman iteration for the primal problem, and the other one is derived from the gradient descent for the dual problem. In addition, the exponential decay of residual and error in corresponding ellipsoidal norms is given. Also we provide an error bound for the result when there is noise in the measurements.

If some or all elements of the matrix A are negative, we can also obtain the same results, except that in the definition of $A_{i,+}$ and $A_{+,j}$, we have to use the summation of the absolute value of elements in rows and columns, which is the l_1 norm of the rows and columns. Numerical results in computed tomography have been presented, to better illustrate the theoretical results.

Acknowledgement

The author would like to thank Luminita 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), 50-55.

[2] A. Andersen and A. Kak, "Simultaneous Algebraic Reconstruction Technique (SART): a Superior Implementation of the ART Algorithm", *Ultrasonic Imaging*, 6 (1984), 81-94.

[3] C. Atkinson and J. Soria, "An Efficient Simultaneous Reconstruction Technique for Tomographic Particle Image Velocimetry", *Experiments in Fluids*, 47 (2009), 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), 200-217.

[6] 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), 40-58.

[7] 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.

[8] G. Cimmino, "Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari", *La Ricerca Scientifica, II,* 9 (1938), 326-333.

[9] 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), 471-481.

[10] G. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction From Projection*, 2nd edition, Springer, 2009.

[11] M. Jiang and G. Wang, "Convergence Studies on Iterative Algorithms for Image Reconstruction", *IEEE Transactions on Medical Imaging*, 22 (2003), 569-579.

[12] M. Jiang and G. Wang, "Convergence of the Simultaneous Algebraic Reconstruction Technique (SART)," *IEEE Transaction on Image Processing*, 12 (2003), 957-961.

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

[14] L. Landweber, "An Iteration Formula for Fredholm Integral Equations of the First Kind," *The American Journal of Mathematics*, 73 (1951), 615-624.

[15] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.

[16] 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), 460-489.

[17] 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), 93-111.

[18] 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), 72585J.

[19] G. Qu, C. Wang and M. Jiang, "Necessary and Sufficient Convergence Conditions for Algebraic Image Reconstruction Algorithms", *IEEE Transactions on Image Processing*, 18 (2009), 435-440.

[20] L.A. Shepp and B. Logan, "The Fourier Reconstruction of a Head Section", *IEEE Transaction on Nuclear Science*, 21 (1974), 21-34.

[21] L.A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography", *IEEE Transaction on Medical Imaging*, 1 (1982), 113-122.

[22] R. Siddon, "Fast Calculation of the Exact Radiological Path for a Three-Dimensional CT Array," *Medical Physics*, 12 (1986), 252-255.

[23] 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), 5781-5804.

[24] J. Wang and Y. Zheng, "On the Convergence of Generalized Simultaneous Iterative Reconstruction Algorithms", *IEEE Transaction on Image Processing*, 16 (2007), 1-6.

[25] W. Yin, "Analysis and Generalizations of the Linearized Bregman Method", *UCLA CAM Technical Report*, 09-42, 2009.

[26] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman Iterative Algorithms for L1- Minimization with Applications to Compressed Sensing", *Journal on Imaging Sciences*, 1 (2008), 143-168.

[27] 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, to appear.