Convergence of the Simultaneous Algebraic Reconstruction Technique (SART)

Ming Jiang*
Department of Information Science
School of Mathematics, Peking University
Beijing 100871, China.

Ge Wang
Department of Radiology
University of Iowa
Iowa City, IA 52242, USA

October 29, 2001

Abstract

In 1984, the simultaneous algebraic reconstruction technique (SART) was developed as a major refinement of the algebraic reconstruction technique (ART). Since then, the SART approach remains a powerful tool for iterative image reconstruction. However, the convergence of the SART has never been established. In this paper, this long-standing conjecture is proven under the condition that coefficients of the linear imaging system are non-negative. It is shown that from any initial value the sequence generated by the SART converges to a weighted least square solution. The importance of the SART and several relevant issues are also discussed.

1 Introduction

Computed tomography (CT) has been extensively studied for years and widely used. Transmission and emission CT are of our particular interest, both of which are linear inverse problems. Although the filtered back-projection method has been the method of choice by CT manufacturers, efforts are being made to revisit iterative methods [1, 2, 3, 4, 5, 6, 7].

Relative to closed-form solutions such as the filtered back-projection algorithm, the iterative approach has a major potential to achieve a superior performance in handling incomplete, noisy, and dynamic data. Although the iterative approach is generally slow, the computing technology is coming to the point that commercial implementation of iterative methods becomes practical for important radiological applications, including image noise suppression, metal artifact reduction, CT fluoroscopic imaging, and so on. More importantly, the theoretical findings accumulated over the past three decades have greatly improved our knowledge of the iterative methods, and formed a solid foundation for further advancement.

Historically, the algebraic reconstruction technique (ART) was the first iterative algorithm applied in CT [8]. In 1984, the simultaneous algebraic reconstruction technique (SART) was proposed as a major refinement of the ART [9]. Since then, the SART remains a powerful tool for iterative reconstruction. Recently, the SART has been used in a number of studies with impressive results [10, 11, 12, 13, 14]. However, we notice that the convergence of the SART has never been proved.

In the following, we will prove that the SART does converge, assuming that coefficients of the linear imaging system are non-negative, as in the cases of transmission and emission CT. In § 2, we will review the SART, describe our assumptions, and define some notations. In § 3, we will establish several properties of the SART. In § 4, we will prove the convergence of the SART.

2 Preliminaries

We model a linear imaging system as follows:

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

where $A = (A_{i,j})$ denotes an $M \times N$ matrix, $b = (b_1, \cdots, b_M)^{\mathrm{tr}} \in \mathbf{R}^M$ the observed data, $x = (x_1, \cdots, x_N)^{\mathrm{tr}} \in \mathbf{R}^N$ an underlying image. We use $^{\mathrm{tr}}$ for the transpose of a vector/matrix. We define

$$A_{i,+} = \sum_{j=1}^{N} A_{i,j} \quad \text{for } i = 1, \cdots, M, \tag{2}$$

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

$$\bar{b}(x) = Ax. \tag{4}$$

*Currently, Ming Jiang is a visiting professor with Micro-CT Laboratory, Department of Radiology, University of Iowa, Iowa City, IA 52242, USA.

0-7803-7147-X/01/$10.00©2001 IEEE

The SART formula [9] is expressed as

$$x_j^{(k+1)} = x_j^{(k)} + \frac{1}{A_{+,j}} \sum_{i=1}^{M} \frac{A_{i,j}}{A_{i,+}} \left(b_i - \bar{b}_i(x^{(k)}) \right) \quad (5)$$

for $k = 0, 1, \cdots$. A generalized version was proposed in [15] as

$$x_j^{(k+1)} = x_j^{(k)} + \frac{\omega}{A_{+,j}} \sum_{i=1}^{M} \frac{A_{i,j}}{A_{i,+}} \left(b_i - \bar{b}_i(x^{(k)}) \right) \quad (6)$$

for $k = 0, 1, \cdots$, where ω denotes a relaxation parameter in $(0, 2)$.

Our assumptions are as follows.

- $A_{i,j} \geq 0$, for $i = 1, \cdots, M$ and $j = 1, \cdots, N$.
- $A_{+,j} \neq 0$ and $A_{i,+} \neq 0$ for $j = 1, \cdots, N$ and $i = 1, \cdots, M$, respectively.

The first assumption requests that coefficients of the imaging system be nonnegative, which is valid in many imaging problems, such as transmission and emission CT. The second assumption appears restrictive, but it is really not. Actually, $A_{+,j} \neq 0$ means that each x_j is measured in some b_i. If $A_{+,j} = 0$, then $A_{i,j} = 0$ for all i, and x_j is un-observable. Hence, that x_j should not be included in the imaging system. On the other hand, $A_{i,+} \neq 0$ means that every b_i carries a certain amount of information about x. If $A_{i,+} = 0$, then $A_{i,j} = 0$ for all j, and no information about x can be inferenced from b_i. Therefore, b_i can be removed from $\mathbf{R}^M$.

The following notations are defined for later use. In either $\mathbf{R}^N$ or $\mathbf{R}^M$, the canonical Euclidean norm is represented by $||\cdot||$, the canonical Euclidean inner product by $< \cdot, \cdot >$, and the zero element by θ. Let V be the diagonal matrix with diagonal element $A_{+,j}$, and W be the diagonal matrix with diagonal element $\frac{1}{A_{i,+}}$. The matrices V and W induce the following inner products on $\mathbf{R}^N$ and $\mathbf{R}^M$:

$$\begin{aligned} < x, x >_V &= < Vx, x >, \quad \text{for } x \in \mathbf{R}^N, \quad (7) \\ < y, y >_W &= < Wy, y >, \quad \text{for } y \in \mathbf{R}^M. \quad (8) \end{aligned}$$

The corresponding V- and W-norms are denoted as $|| \cdot ||_V$ and $|| \cdot ||_W$, respectively. In terms of V and W, the SART formula becomes

$$x^{(k+1)} = x^{(k)} + \omega V^{-1} A^{\mathrm{tr}} W (b - Ax^{(k)}). \quad (9)$$

We use $\mathbf{R}^N$ and $\mathbf{R}^M$ to denote the canonical Euclidean spaces with the canonical Euclidean inner products $< \cdot, \cdot >$, respectively. Also, we have the spaces $\mathcal{X}$ and $\mathcal{Y}$ that are $\mathbf{R}^N$ with the inner product $< \cdot, \cdot >_V$ and $\mathbf{R}^M$ with the inner product $< \cdot, \cdot >_W$, respectively.

3 Properties of the SART

Let $N(A)$ be the null space of A. By the orthogonal decomposition theorem,

$$\mathbf{R}^N = N(A) \oplus N(A)^{\perp}. \quad (10)$$

We use $B^{\perp}$ to denote the orthogonal complement of a subspace B in either $\mathbf{R}^N$ or $\mathbf{R}^M$. Let $R(A^{\mathrm{tr}})$ be the range of A^{tr}. It is well-known that $N(A)^{\perp} = R(A^{\mathrm{tr}})$.

Define the weighted least square functional

$$L(x) = \sum_{i=1}^{M} \frac{1}{A_{i,+}} \left(b_i - \bar{b}_i(x) \right)^2 = ||b - Ax||_W^2 \quad (11)$$

for $x \in \mathbf{R}^N$. It turns out later that the sequence generated by the SART converges to a minimizer of this functional. The gradient of L is

$$\nabla L(x) = -2A^{\mathrm{tr}} W (b - Ax). \quad (12)$$

All the minimizers of L satisfy the following normal equation:

$$A^{\mathrm{tr}} W A x = A^{\mathrm{tr}} W b. \quad (13)$$

Since $< A^{\mathrm{tr}} W A x, x > = ||Ax||_W^2$, it is easy to show that $N(A^{\mathrm{tr}} W A) = N(A)$. Since $A^{\mathrm{tr}} W b \in R(A^{\mathrm{tr}}) = N(A)^{\perp}$, there must exist a solution to (13).

Let S be the set of all solutions to (13). It can be easily verified that $L(x)$ is convex on $\mathbf{R}^N$. Hence, a solution of (13) is also a minimizer of L on $\mathbf{R}^N$, and vice versa. Hence, S is also the set of all minimizers of L on $\mathbf{R}^N$ and a closed convex set of $\mathbf{R}^N$.

Let

$$F(x) = \omega V^{-1} A^{\mathrm{tr}} W (b - Ax). \quad (14)$$

The SART becomes

$$x^{(k+1)} = x^{(k)} + F(x^{(k)}). \quad (15)$$

The descending of L with the SART iteration can be estimated by the following proposition:

Proposition 3.1.

$$L(x^{(k+1)}) - L(x^{(k)}) \leq -\alpha ||x^{(k+1)} - x^{(k)}||_V^2, \quad (16)$$

for $k = 0, 1, \cdots$, *where* $\alpha = \frac{2}{\omega} - 1 > 0$.

Proof. By computation,

$$\begin{aligned} & L(x^{(k+1)}) - L(x^{(k)}) \\ = & \sum_{i=1}^{M} \frac{1}{A_{i,+}} \left(\bar{b}_i(F(x^{(k)})) \right)^2 \\ & - 2 \sum_{i=1}^{M} \frac{1}{A_{i,+}} \left(b_i - \bar{b}_i(x^{(k)}) \right) \cdot \bar{b}_i(F(x^{(k)})). \end{aligned}$$

For the first term, by the convexity of the function $t \mapsto t^2$, we have

$$\begin{aligned}
&\sum_{i=1}^{M} \frac{1}{A_{i,+}} \left(\bar{b}_i(F(x^{(k)}))\right)^2 \\
=&\sum_{i=1}^{M} A_{i,+} \left(\sum_{j=1}^{N} \frac{A_{i,j}}{A_{i,+}} F_j(x^{(k)})\right)^2 \\
\leq&\sum_{j=1}^{N} A_{+,j} F_j(x^{(k)})^2 = ||x^{(k+1)} - x^{(k)}||_V^2.
\end{aligned}$$

For the second term,

$$\begin{aligned}
&2\sum_{i=1}^{M} \frac{1}{A_{i,+}} \left(b_i - \bar{b}_i(x^{(k)})\right) \cdot \bar{b}_i(F(x^{(k)})) \\
=&2\sum_{j=1}^{N} F_j(x^{(k)}) \sum_{i=1}^{M} \frac{A_{i,j}}{A_{i,+}} \left(b_i - \bar{b}_i(x^{(k)})\right) \\
=&\frac{2}{\omega} \sum_{j=1}^{N} A_{+,j} \left[F_j(x^{(k)})\right]^2 = \frac{2}{\omega} ||x^{(k+1)} - x^{(k)}||_V^2.
\end{aligned}$$

Therefore, the conclusion follows immediately. □

Corollary 3.2.

$$L(x^{(k+1)}) + \alpha \sum_{j=0}^{k} ||x^{(j+1)} - x^{(j)}||_V^2 \leq L(x_0). \quad (17)$$

And the series $\sum_{j=0}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2$ *is convergent.*

Proof. Note

$$\begin{aligned}
L(x^{(1)}) + \alpha ||x^{(1)} - x^{(0)}||_V^2 \leq & \quad L(x^{(0)}), \\
& \quad \vdots \\
L(x^{(k+1)}) + \alpha ||x^{(k+1)} - x^{(k)}||_V^2 \leq & \quad L(x^{(k)}).
\end{aligned}$$

Summing up all these inequalities, we have

$$L(x^{(k+1)}) + \alpha \sum_{j=0}^{k} ||x^{(j+1)} - x^{(j)}||_V^2 \leq L(x^{(0)}). \quad (18)$$

Since $L(x) \geq 0$, the convergence of the series is obvious. □

4 Convergence of the SART

Now consider the Hilbert space $\mathcal{X}$. By orthogonal decomposition, we have

$$\mathcal{X} = N(A) \oplus N(A)^{\perp_V}, \quad (19)$$

where $B^{\perp_V}$ denotes the orthogonal complement of a subspace B in $\mathcal{X}$. Since S is also a closed convex set in $\mathcal{X}$, it must contain a unique element x^* with the minimal V-norm. Clearly, $x^* \in N(A)^{\perp_V}$.

By the normal equation (13), the SART can be written as

$$x^{(k+1)} - x^* = x^{(k)} - x^* + \omega V^{-1} A^{\mathrm{tr}} W A (x^* - x^{(k)}). \quad (20)$$

Let $z^{(k)} = x^{(k)} - x^*$, we have

$$z^{(k+1)} = z^{(k)} - \omega V^{-1} A^{\mathrm{tr}} W A z^{(k)}. \quad (21)$$

The inner product of (21) with $z^{(k)}$ in $\mathcal{X}$ is

$$< V z^{(k+1)}, z^{(k)} >= ||z^{(k)}||_V^2 - \omega ||A z^{(k)}||_W^2.$$

The left hand is, noting that V is diagonal,

$$\begin{aligned}
& < V z^{(k+1)}, z^{(k)} > \\
=& < V z^{(k+1)}, z^{(k+1)} > + < V z^{(k+1)}, z^{(k)} - z^{(k+1)} > \\
=& ||z^{(k+1)}||_V^2 - ||z^{(k)} - z^{(k+1)}||_V^2 \\
& \qquad + < V z^{(k)}, z^{(k)} - z^{(k+1)} > \\
=& ||z^{(k+1)}||_V^2 - ||z^{(k)} - z^{(k+1)}||_V^2 \\
& \qquad + \omega < V z^{(k)}, V^{-1} A^{\mathrm{tr}} W A z^{(k)} > \\
=& ||z^{(k+1)}||_V^2 - ||z^{(k)} - z^{(k+1)}||_V^2 \\
& \qquad + \omega < z^{(k)}, A^{\mathrm{tr}} W A z^{(k)} > \\
=& ||z^{(k+1)}||_V^2 - ||z^{(k)} - z^{(k+1)}||_V^2 + \omega ||A z^{(k)}||_W^2.
\end{aligned}$$

Therefore,

$$||z^{(k+1)}||_V^2 = ||z^k||_V^2 + ||z^{(k)} - z^{(k+1)}||_V^2 - 2\omega ||A z^{(k)}||_W^2, \quad (22)$$

especially,

$$||z^{(k+1)}||_V^2 \leq ||z^k||_V^2 + ||z^{(k)} - z^{(k+1)}||_V^2. \quad (23)$$

Proposition 4.1. *The sequence* $\{z^k\}$ *is bounded.*

Proof. By (23) and using the same method as in the proof of Corollary 3.2, we have

$$||z^{(k+1)}||_V^2 \leq ||z^{(0)}||_V^2 + \sum_{j=0}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2$$

Since the last series is convergent by Corollary 3.2, the conclusion follows immediately. □

Let $r_k = ||z^{(k)}||_V^2$.

Proposition 4.2. *The sequence* $\{r_k\}$ *is convergent, i.e., the limit* $\lim_{k \to \infty} r_k$ *exists and is finite.*

Proof. Since $\{r_k\}$ is bounded, both the limit superior $\hat{r} = \limsup_{k\to\infty} r_k$ and limit inferior $\check{r} = \liminf_{k\to\infty} r_k$ exist. We can show that $\hat{r} = \check{r}$ as follows. For any $p > q$, we have

$$r_p \le r_q + \sum_{j=q}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2.$$

Fixing q and taking the limit superior with respect to p, we obtain

$$\hat{r} \le r_q + \sum_{j=q}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2.$$

Then, taking the limit inferior with respect to q,

$$\hat{r} \le \check{r} + \lim_{q\to\infty} \sum_{j=q}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2.$$

Because $\sum_{j=0}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2$ exists,

$$\lim_{q\to\infty} \sum_{j=q}^{\infty} ||x^{(j+1)} - x^{(j)}||_V^2 = 0.$$

Therefore, $\hat{r} = \check{r}$. □

Proposition 4.3. *The following series is convergent:*

$$\sum_{k=0}^{\infty} ||Az^{(k)}||_W^2. \tag{24}$$

Hence $||Az^{(k)}||_W \to 0$.

Proof. By (22),

$$||Az^{(k)}||_W^2 = \frac{1}{2\omega}\left(r_k - r_{k+1} + ||z^{(k)} - z^{(k+1)}||_V^2\right)$$

The conclusion follows from Corollary 3.2 and Proposition 4.2. □

Before we prove the main theorem on the convergence of the SART, we need the following lemma:

Lemma 4.4. *There exists a positive constant λ such that*

$$||Av||_W \ge \lambda ||v||_V \quad \textit{for } v \in N(A)^{\perp_V}. \tag{25}$$

Proof. We will prove the inequality by contradiction. If the inequality is false, there must exist $v_n \in N(A)^{\perp_V}$ such that

$$||Av_n||_W \le \frac{1}{n}||v_n||_V, \quad \text{for } n = 1, \cdots$$

Dividing both sides by $||v_n||_V$ and setting $u_n = \frac{v_n}{||v_n||_V}$, we have

$$||Au_n||_W \le \frac{1}{n}. \tag{26}$$

Note that $u_n \in N(A)^{\perp_V}$ and $||u_n||_V = 1$. Because $\{u_n\}$ is bounded, it has a convergent subsequence. Let it be $\{u_{n_j}\}$ and assume that $u_{n_j} \longrightarrow u_0$. Then we have $||u_0||_V = 1$. By closedness of $N(A)^{\perp_V}$, $u_0 \in N(A)^{\perp_V}$. By (26), we have $||Au_0||_W = 0$. Hence, $Au_0 = \theta$, which implies that $u_0 \in N(A)$. Then, $u_0 \in N(A) \cap N(A)^{\perp_V} = \{\theta\}$. Therefore, $u_0 = \theta$, which is in contradiction with $||u_0||_V = 1$. □

Let P and Q be the orthogonal projection from $\mathcal{X}$ to $N(A)$ and $N(A)^{\perp_V}$ respectively, we are now ready to prove the following main theorem.

Theorem 4.5. *The sequence $x^{(k)}$ generated by the SART (6) converges to $P[x_0] + x^* \in S$, for any $\omega \in (0, 2)$, where x^* is the solution of the normal equation (13) with the minimum V-norm. The limit is a solution to the normal equation (13), hence a global minimizer of L.*

Proof. For any $u \in N(A)$ and any $x \in \mathbf{R}^N$,

$$\begin{aligned} < F(x), u >_V &=< VF(x), u >=< \omega A^{\mathrm{tr}} W(b - Ax), u > \\ &= \omega < W(b - Ax), Au >= 0. \end{aligned}$$

Therefore, $F(x) \in N(A)^{\perp_V}$. By (15), we have $P[x^{(k+1)}] = P[x^{(k)}] + P[F(x^{(k)})] = P[x^{(k)}]$, for $k = 0, \cdots$. Hence, $P[x^{(k+1)}] = P[x^{(0)}]$. By construction, $x^* \in N(A)^{\perp_V}$. Therefore,

$$Az^{(k)} = AQ[x^{(k+1)} - x^*],$$

noting $AP[x^{(k+1)}] = \theta$, $Q[x^*] = x^*$ and $x^{(k+1)} = P[x^{(k+1)}] + Q[x^{(k+1)}]$. Since $Q[x^{(k+1)} - x^*] \in N(A)^{\perp_V}$, by Lemma 4.4,

$$||Az^{(k)}||_W \ge \lambda ||Q[x^{(k+1)} - x^*]||_V.$$

By Proposition 4.3, $Q[x^{(k+1)}] \longrightarrow Q[x^*] = x^*$. Using $x^{(k+1)} = P[x^{(k+1)}] + Q[x^{(k+1)}]$ again, the conclusion follows immediately. □

5 Discussion and Conclusion

The convergence of the SART can also be proved using a general theorem of Bialy [16], which involves more advanced mathematics on spectrum theory. We have decided to present the above relatively elementary proof for easy understanding. Also, when the coefficients of the imaging system take negative values, $A_{i,+}$ may be negative, the weighted least square functional $L(x)$ is non-convex and ill-defined. Hence the SART must be modified. The authors have successfully found the modified version together with a proof of convergence, but it is beyond the scope of this paper and will be published somewhere. Further research topics include regularization, acceleration, extension and evaluation of the SART, as well as the relationship among the SART and other iterative formulas.

In conclusion, we have proved the long-standing conjecture on the convergence of the SART under the condition that coefficients of the linear imaging system are non-negative. It is shown that the sequence generated by the SART converges to a minimizer of a weighted least square functional from any initial guess in the real space.

Acknowledgments

This work was supported in part by the National Institutes of Health (NIDCD R01 DC00581). Also, Dr. G. Wang thanks Drs. Y. Censor, W. Han, G. Herman, Z. Li, Y. Li, and D. L. Snyder for discussions. Especially, Dr. Censor confirmed the unavailability of a proof-of-convergence for the SART.

References

[1] Y. Censor, "Finiteseries-expansion reconstruction methods," *Proc. IEEE*, vol. 71, pp. 409 – 419, 1983.

[2] Y. Censor, "On some optimization techniques in image reconstruction from projections," *Applied Numerical Mathematics*, vol. 3, pp. 365 – 391, 1987.

[3] M. Hanke and P. C. Hansen, "Regularization methods for large scale problems," *Surveys Math. Indust.*, vol. 3, pp. 235 – 315, 1993.

[4] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, 1997.

[5] M. Bertero and P. Boccacci, *Inverse Problems in Imaging*, Institute of Physical Publishing, Bristol and Philadelphis, 1998.

[6] G. Wang, M. W. Vannier, and P. C. Cheng, "Iterative X-ray cone-beam tomography for metal artifact reduction and local region reconstruction," *Microscopy and Microanalysis*, vol. 5, pp. 58 – 65, 1999.

[7] Yousef Saad and Henk A. van der Vorst, "Iterative solution of linear systems in the 20th century," *Journal of Computational and Applied Mathematics*, vol. 123, pp. 1 – 33, 2000.

[8] 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*, vol. 29, pp. 471 – 482, 1970.

[9] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, pp. 81 – 94, 1984.

[10] G. Wang, D. L. Snyder, J. A. O'Sullivan, and M. W. Vannier, "Iterative deblurring for CT metal artifact reduction," *IEEE Trans. Medical Imaging*, vol. 15, pp. 657–664, 1996.

[11] G. Wang, M. W. Vannier, M. W. Skinner, M. G. P. Cavalcanti, and G. Harding, "Spiral CT image deblurring for cochlear implantation," *IEEE Trans. Medical Imaging*, vol. 17, pp. 251 – 262, 1998.

[12] G. Wang, G. D. Schweiger, and M. W. Vannier, "An iterative algorithm for X-ray CT fluoroscopy," *IEEE Trans. Medical Imaging*, vol. 17, pp. 853–856, 1998.

[13] K. Mueller and R. Yagel et al., "Anti-aliased three-dimensional cone-beam reconstruction of low-contrast objects with algebraic methods," *IEEE Trans. Medical Imaging*, vol. 18, no. 6, pp. 519 – 537, 1999.

[14] K. Mueller and R. Yagel et al., "Fast implementations of algebraic methods for three-dimensional reconstruction from cone-beam data," *IEEE Trans. Medical Imaging*, vol. 18, no. 6, pp. 538 – 548, 1999.

[15] A. H. Andersen, "Algebraic reconstruction in CT from limited views," *IEEE Trans. Medical Imaging*, vol. 8, pp. 50 – 55, 1989.

[16] Horst Bialy, "Iterative behandlung linearer funktionalgleichung," *Arch. for Rational Mechanics*, vol. 4, pp. 166 – 176, 1959.