Projected non-stationary simultaneous iterative methods

Touraj Nikazad*, Mahdi Mirzapour

*School of of Mathematics, Iran University of Science and Technology, Tehran 16846-13114, Iran

(Communicated by M. Eshaghi)

Abstract

In this paper, we study Projected non-stationary Simultaneous Iterative Reconstruction Techniques (P-SIRT). Based on algorithmic operators, convergence result are adjusted with Opial’s Theorem. The advantages of P-SIRT are demonstrated on examples taken from tomographic imaging.

Keywords: simultaneous iterative reconstruction techniques; convex feasibility problem; (firmly) nonexpansive operator; cutter operator.

2010 MSC: Primary 65J22, 65J15; Secondary 65F10.

1. Introduction

Large-scale discretizations of ill-posed problems (as imaging problems in tomography) lead to large, sparse and ill-posed (sensitive to noise) linear systems of equations (which may be inconsistent) of the form

\[ Ax = b. \]  

(1.1)

Many problems as image reconstruction [30, 12, 13, 26, 24, 23], computed tomography [21, 22, 34], image recovery [33, 35], image restoration [36], image registration [29], seismic imaging [20], image fusion [17], radar imaging [14] lead to a linear system as (1.1).

Finding \( x^* \in \mathbb{R}^n \) such that \( Ax^* = b \) is a special case of convex feasibility problems (CFPs). Actually many problems in mathematics and physical sciences can be modeled as a CFP, i.e., a problem of finding a point \( x \in Q = \bigcap_{i=1}^{m} Q_i \) where \( \{Q_i\}_{i=1}^{m} \subseteq \mathbb{R}^n \) are closed convex sets. Using fixed point iterative methods based on algorithmic operators has been suggested by many researchers for solving CFPs, see, e.g., [2, 8]. One of the most important class of algorithmic operators is
projection algorithms that play a main role in the area of constructive solution of CFPs. Projection algorithms are iterative algorithms that use projections onto sets. We next give some instances of such algorithms.

Let \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \). Let \( a^i \) and \( b_i \) denote the \( i \)-row of \( A \) and \( b \), respectively. Therefore projection of \( x \in \mathbb{R}^n \) onto the \( i \)-hyperplane, i.e. \( H_i = \{ x \in \mathbb{R}^n | \langle a^i, x \rangle = b_i \} \), is

\[
P_{H_i}(x) = x + \frac{b_i - \langle a^i, x \rangle}{\|a^i\|^2} a^i \quad i = 1, 2, \ldots, m
\]  

(1.2)

where \( i = 1, 2, \ldots, m \). To simplifying the notation we denote \( P_{H_i}(x) = P_i(x) \).

The projection operators can be used in various ways. We briefly explain a special case of sequential and simultaneous methods which use two different ways of projection operators. Algebraic Reconstruction Technique (ART) \[21\] is a sequential method which executes as follows. Let \( x^0 \in \mathbb{R}^n \) be an arbitrary starting point. The ART algorithm projects the current iteration \( x^k \) onto a hyperplane, e.g. \( H_i \), and puts \( x^{k+1} = P_i(x^k) \). Let \( T = P_m \cdots P_2 P_1 \) where \( P_i \) is defined in (1.2). One cycle of the ART method is performed by acting \( T \) on the staring point. In this way, we obtain a sequence of cycles which is a subsequence of iterations, see Figure 1. Simultaneous algorithms project \( x^k \) onto all hyperplanes \( \{H_i\}_{i=1}^m \) simultaneously. The next iteration is performed as convex combination of \( m \) new projected points, see Figure 2.

We now explain this algorithm with more details. Let \( T = \sum_{i=1}^m \omega_i P_i \), where \( \sum_{i=1}^m \omega_i = 1 \) and \( \omega_i \geq 0 \).

Using (1.2) we get

\[
T(x) = \sum_{i=1}^m \omega_i P_i
\]

(1.3)

\[
= \sum_{i=1}^m \omega_i x + \frac{b_i - \langle a^i, x \rangle}{\|a^i\|^2} a^i \\
= x + A^T M (b - Ax)
\]

(1.4)

where

\[
M = \text{diag} \left( \frac{\omega_1}{\|a^1\|^2}, \ldots, \frac{\omega_m}{\|a^m\|^2} \right).
\]
Therefore, using (1.3) and (1.4), the fixed point iterative method \( x^{k+1} = T(x^k) \) is a special case of SIRT. In general, the SIRT is defined as the following iteration algorithm

\[
x^{k+1} = x^k + \lambda_k S A^T M (b - Ax^k) \quad k = 0, 1, 2, \ldots
\]

where \( \lambda_k \in [\epsilon, \frac{2-\tau}{\sigma_1}] \) are relaxation parameters and \( \sigma_1 \) is the largest singular value of \( M^{\frac{1}{2}} A S^{\frac{1}{2}} \). Also, \( M \) and \( S \) are assumed symmetric positive definite matrices. Several well-known fully simultaneous methods can be written in the form of (1.5) for appropriate choices of \( M \) and \( S \) matrices. We below give some instances

- Landweber’s method [28]: \( S = M = I \),
- Cimmino’s method [15]: \( S = I \) and \( M = \frac{1}{m} \text{diag}\left(\frac{1}{\|a_i\|^2}\right) \),
- CAV (Component Averaging method) method [? 12]: \( S = I \) and \( M = \text{diag}(1/\sum_{j=1}^n N_j a_{ij}^2) \) where \( N_j \) is the number of non-zeroes in the \( j \)th column of \( A \),
- DROP (Diagonally Relaxed Orthogonal Projection) method [? ]: \( S = \text{diag}(m/N_j) \) and \( M \) any symmetric positive definite matrix.

Furthermore, the SART method [11] and the symmetric Kaczmarz’s method [6] can be rewritten as (1.5).

When solving an inverse problem, the use of constraints (like nonnegativity) and prior information are well known techniques to improve the quality of the obtained solution because incorporate prior physical knowledge about the solution leads to smaller reconstruction errors, see [5, 4, 3, 7, 25, 32, 37].

In this paper we consider the projected version of equation (1.5) in a finite dimensional Euclidean space \( \mathbb{R}^n \). Let \( C \subseteq \mathbb{R}^n \) denote a closed convex set and \( P_C \) be the metric projection onto \( C \). Assume that \( \{\lambda_k\}_{k=0}^{\infty} \) is a sequence of positive relaxation parameters. Now consider the following algorithm.

**Algorithm 1.1. (P-SIRT)**

Initialization: \( x^0 \in \mathbb{R}^n \) is arbitrary.

Iterative Step: given \( x^k \), compute

\[
x^{k+1} = P_C \left( x^k + \lambda_k S A^T M (b - Ax^k) \right) \quad k = 0, 1, 2, \ldots
\]
In Next section, using algorithmic operators, we adjust Algorithm 1.1 with a corollary of generalization of Opial’s Theorem, where relaxation parameters are changed in each iteration. The paper is organized as follows. In section 2, we recall some definitions and properties of some algorithmic operators and give the convergence analysis of Algorithm 1.1. At the end, the capability of the main result is examined in section 3 using some numerical tests from the medical imaging field.

2. Preliminaries and Notations

Throughout this section, we consider \( T : H \rightarrow H \) with nonempty fixed point set, i.e., \( \text{Fix} T \neq \emptyset \) where \( H \) is a Hilbert space and \( Id \) denotes the identity operator on \( H \). The following definitions, taken from [8], will be useful in our future analysis.

Definition 2.1. Let \( T : H \rightarrow H \) and \( \alpha \in [0, 2] \). The operator \( T_\alpha \) defined by

\[
T_\alpha := (1 - \alpha)Id + \alpha T
\]  

(2.1)
is called an \( \alpha \)-relaxation or, shortly, relaxation of the operator \( T \). If \( \alpha \in (0, 2) \), then \( T_\alpha \) is called a strictly (or strict) relaxation of \( T \).

Definition 2.2. We say that an operator \( T : H \rightarrow H \) is nonexpansive (NE), if

\[
\|T(x) - T(y)\| \leq \|x - y\|
\]  

(2.2)

for all \( x, y \in H \). Also \( T \) is an \( \alpha \)-contraction, where \( \alpha \in (0, 1) \) or, shortly, a contraction if

\[
\|T(x) - T(y)\| \leq \alpha \|x - y\|
\]  

(2.3)

for all \( x, y \in H \).

Another useful class of operators is the class of cutter operators, namely

Definition 2.3. An operator \( T : H \rightarrow H \) with nonempty fixed point set is called cutter if

\[
\langle x - T(x), z - T(x) \rangle \leq 0
\]  

(2.4)

for all \( x \in H \) and \( z \in \text{Fix} T \).

Remark 2.4. Based on [8] Remark 2.1.31] the operator \( T \) is a cutter if and only if

\[
\langle T(x) - x, z - x \rangle \geq \|T(x) - x\|^2
\]  

(2.5)

for all \( x \in H \) and \( z \in \text{Fix} T \).

Definition 2.5. We say that an operator \( T : H \rightarrow H \) is firmly nonexpansive (FNE), if

\[
\langle T(x) - T(y), x - y \rangle \geq \|T(x) - T(y)\|^2
\]  

(2.6)

for all \( x, y \in H \).

Based on [8] Remark 2.1.31], an \( \alpha \)-relaxed cutter operator is defined as follows.
**Definition 2.6.** Let \( T : H \to H \) has a fixed point. Then the operator \( T \) is an \( \alpha \)-relaxed cutter, or, shortly, relaxed cutter where \( \alpha \in [0, 2] \), if
\[
\langle T_\alpha(x) - x, z - x \rangle = \alpha \langle T(x) - x, z - x \rangle \geq \|T(x) - x\|^2
\]
for all \( x \in H \) and \( z \in \text{Fix}T \). If \( \alpha \in (0, 2] \), then \( T_\alpha \) is called a strictly relaxed cutter operator of \( T \).

**Definition 2.7.** Let \( \alpha \geq 0 \) and assume that \( T : H \to H \) has a fixed point. We say that \( T \) is \( \alpha \)-strongly quasi-nonexpansive (\( \alpha \)-SQNE), if
\[
\|T(x) - z\|^2 \leq \|x - z\|^2 - 2\alpha \|T(x) - x\|^2
\]
for all \( x \in H \) and \( z \in \text{Fix}T \). Also, the operator \( T \) satisfying (2.8) with \( \alpha > 0 \) is called strongly quasi-nonexpansive (SQNE) operator.

Following theorem presents the relationship between strictly relaxed cutter and SQNE operators.

**Theorem 2.8.** [8, Theorem 2.1.39 and Corollary 2.1.40] Assume that \( T : H \to H \) has a fixed point and let \( \lambda \in (0, 2] \). Then \( T \) is a \( \lambda \)-relaxed cutter if and only if \( T \) is \( 2 - \lambda \)-SQNE, i.e.,
\[
\|T_\lambda(x) - z\|^2 \leq \|x - z\|^2 - \frac{2 - \lambda}{\lambda} \|T_\lambda(x) - x\|^2
\]
for all \( x \in H \) and all \( z \in \text{Fix}T \).

**Definition 2.9.** An operator \( T : H \to H \) is demi-closed at 0 if for any weakly converging sequence \( x_k \rightharpoonup y \in H \) with \( T(x_k) \to 0 \) we have \( T(y) = 0 \).

**Remark 2.10.** It is well known, see [31, Lemma 2], the operator \( T - Id \) is demi-closed at 0 where \( T : H \to H \) is a nonexpansive operator.

We now verify, using [9 Corollary 9.14.], that the sequence generated by Algorithm (1.1) converges.

**Corollary 2.11.** [9 Corollary 9.14.] and [8 Corollary 3.7.3] Let \( T : H \to H \) be a cutter operator (e.g., a firmly nonexpansive operator having a fixed point) and \( x^0 \in H \) is an arbitrary point. Assume that the sequence \( \{x^k\}_{k=0}^\infty \) is generated by
\[
x^{k+1} = P_C \left( x^k + \lambda_k \left( T(x^k) - x^k \right) \right) \quad \text{for} \quad k = 1, 2, \ldots
\]
where \( \lambda_k \in (0, 2) \).

(i) If \( \liminf_{k \to \infty} \lambda_k (2 - \lambda_k) > 0 \), then \( \{x^k\}_{k=0}^\infty \) converges weakly to a fixed point of \( T \).

(ii) If \( H \) is finite-dimensional and \( \sum_{k=0}^\infty \lambda_k (2 - \lambda_k) = \infty \), then \( \{x^k\}_{k=0}^\infty \) converges to a fixed point of \( T \).

Let \( B = S^{\frac{1}{2}}A^T M A S^{\frac{1}{2}} = (M^{\frac{1}{2}} A S^{\frac{1}{2}})^T (M^{\frac{1}{2}} A S^{\frac{1}{2}}) \) then the spectral radius of \( B \) is denoted by \( \rho(B) = \sigma_1^2 \) where \( \sigma_1 \) is the largest singular value of \( M^{\frac{1}{2}} A S^{\frac{1}{2}} \). We next present a useful lemma from [18].
Lemma 2.12. [18 Lemma 3.1] Let $q = \|I - \lambda B\|$. Assume that $\text{rank}(A) = n$ and $\sigma_1 > \sqrt{2}\sigma_n$. Further assume that $\lambda$ fulfills $0 < \epsilon \leq \lambda \leq (2 - \epsilon)/\sigma_1^2$. Then

$$q = \begin{cases} 1 - \lambda \sigma_n^2, & 0 < \lambda \leq \frac{2}{\sigma_1^2 + \sigma_n^2} \\ \lambda \sigma_1^2 - 1, & \frac{2}{\sigma_1^2 + \sigma_n^2} \leq \lambda < \frac{2}{\sigma_1^2}. \end{cases} \tag{2.11}$$

Remark 2.13. It should be noted that for inverse problems $\sigma_n \ll \sigma_1$ and hence $2/(\sigma_1^2 + \sigma_n^2) \approx 2/\sigma_1^2$. Therefore, we will consider only the case $q = 1 - \lambda \sigma_n^2$. Furthermore, one can avoid the assumption $\text{rank}(A) = n$ and consider the rank-deficient case using [18 Lemma 3.9].

We next present the convergence analysis of Algorithm 1.1.

Theorem 2.14. The sequence generated by Algorithm 1.1, where $\lambda_k \in [\epsilon, \frac{2}{\sigma_1^2}]$, converges to a solution $x^\ast$ of $\min \|Ax - b\|_M$.

Proof. Since $\lambda_k \in [\epsilon, \frac{2}{\sigma_1^2}]$ we can rewrite the Algorithm 1.1 as below

$$x^{k+1} = U(x^k) = P_C \left( x^k + \frac{\lambda_k}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} SATM(b - Ax^k) \right) \tag{2.12}$$

$$= P_C \left( x^k + \lambda_k \left( T(x^k) - x^k \right) \right) \tag{2.13}$$

$$= P_C T_{\lambda_k}(x^k) \tag{2.14}$$

where

$$T(x) = x + \frac{1}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} SATM(b - Ax). \tag{2.15}$$

Furthermore, we have

$$\|T(x) - T(y)\| = \left\| (x - y) - \frac{1}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} SATMA(x - y) \right\|$$

$$= \left\| \left( I - \frac{1}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} SATMA \right) (x - y) \right\|$$

$$= \left\| S^{\frac{1}{2}} \left( I - \frac{1}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} S^{\frac{1}{2}}ATMAS^{\frac{1}{2}} \right) S^{\frac{1}{2}} (x - y) \right\|$$

$$\leq \left\| S^{\frac{1}{2}} \left( I - \frac{1}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} S^{\frac{1}{2}}ATMAS^{\frac{1}{2}} \right) \right\| (x - y) \right\|$$

$$= \left\| \left( I - \frac{1}{\rho(S^{\frac{1}{2}}ATMAS^{\frac{1}{2}})} S^{\frac{1}{2}}ATMAS^{\frac{1}{2}} \right) \right\| (x - y) \right\|.$$

Based on Lemma 2.12 Remark 2.13 and setting $\tilde{A} = S^{\frac{1}{2}}A$, we have

$$\alpha = \|I - \frac{1}{\rho(\tilde{A}^T \tilde{A})} \tilde{A}^T \tilde{MA}\| = 1 - \sigma_n^2 < 1.$$

Thus operator $T$ is an $\alpha$-contraction operator. Also based on [8 Theorem 2.2.34], $T$ is a $(1 + \alpha)$-relaxed firmly nonexpansive operator. Using [8 Corollary 2.2.11] $T$ is a firmly nonexpansive and
consequently based on the first part of [8, Theorem 2.2.5] \( T \) is a cutter operator. Using Remark 2.10 we know that the operator \( T - I \) is demi-closed at 0. Therefore based on Corollary 2.11 the sequence \( \{x^k\} \) converges weakly to a fixed point of \( T \). Since we are using finite dimensional space \( \mathbb{R}^n \) we obtain \( x^k \to x^* \) such that \( T(x^*) = x^* \). It gives \( A^T M (b - Ax^*) = 0 \) which is equivalent to the fact that \( x^* \) is a minimizer of \( \| Ax - b \|_M \). □

3. Numerical Result

In this section we report some numerical results in field of medical imaging. Our numerical results show the effect of using projection operator after each iteration. Furthermore we suggest a rule for picking relaxation parameters.

In following two tables we show error histories for Landweber, Cimmino, CAV and DROP algorithms without constraint (\( C = \mathbb{R}^n \)), with non-negativity constraints (\( C = \mathbb{R}^+_n \)), and with box constraints (\( C = [0, 1]^n \)) within 40 iterations. For all of algorithms, we use the following strategy for picking relaxation parameters that were proposed in [18, 19].

\[
\lambda_k = \begin{cases} 
\sqrt{2\sigma_1^{-2}} & \text{for } k = 0, 1 \\
2\sigma_1^{-2} \frac{1 - \zeta_k}{(1 - \zeta_k^2)^2} & \text{for } k \geq 0
\end{cases} 
\] (3.1)

where \( \sigma_1 \) is largest singular value of \( M^{1/2} A S^{1/2} \) and \( \zeta_k \) are roots of a certain polynomial such that \( 0 < \zeta_k < \zeta_{k+1} \) and \( \lim_{k \to \infty} \zeta_k = 1 \).

The test is taken from the field of image reconstruction from projections using the SNARK93 software package [27]. We work with the standard head phantom from [21]. The phantom is discretized into \( 63 \times 63 \) pixels, and 16 projections (evenly distributed between 0 and 174 degrees) with 99 rays per projection are used. The resulting matrix \( A \) has dimension \( 1584 \times 3969 \), so that the system of equations is highly underdetermined. In addition to \( A \), the software also produces a noise-free right-hand side \( b_{\text{snark}} \) and a phantom (translated into vector form) \( x^* \). Using SNARK93’s right-hand side \( b_{\text{snark}} \), which is not generated as the product \( Ax^* \), we avoid committing an inverse crime where the exact same model is used in the forward and reconstruction models. Apart from using noise-free data we also added additive independent Gaussian noise of mean 0 and relative noise-level (\( \| \delta b \| / \| b_{\text{snark}} \| \)) 5% where \( b_{\text{noisy}} = b_{\text{snark}} + \delta b \).

Table 1: The smallest relative error with noiseless (top) and noisy data (down) using Algorithm 1.1

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( C = \mathbb{R}^n )</th>
<th>( C = \mathbb{R}^+_n )</th>
<th>( C = [0, 1]^n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landweber</td>
<td>0.2623</td>
<td>0.2571</td>
<td>0.2571</td>
</tr>
<tr>
<td>Cimmino</td>
<td>0.2338</td>
<td>0.2218</td>
<td>0.2218</td>
</tr>
<tr>
<td>CAV</td>
<td>0.2207</td>
<td>0.2014</td>
<td>0.2014</td>
</tr>
<tr>
<td>DROP</td>
<td>0.2379</td>
<td>0.2379</td>
<td>0.2379</td>
</tr>
<tr>
<td>Landweber</td>
<td>0.2713</td>
<td>0.2621</td>
<td>0.2621</td>
</tr>
<tr>
<td>Cimmino</td>
<td>0.2686</td>
<td>0.2316</td>
<td>0.2316</td>
</tr>
<tr>
<td>CAV</td>
<td>0.2665</td>
<td>0.2157</td>
<td>0.2157</td>
</tr>
<tr>
<td>DROP</td>
<td>0.2665</td>
<td>0.2157</td>
<td>0.2157</td>
</tr>
</tbody>
</table>

In second test we give a strategy for picking relaxation parameters. Assume that the linear system (1.1) is consistent. This strategy is based on picking \( \lambda_k \) such that the error \( \| x^k - x^* \| \) is minimized.
in each iteration where \(x^*\) is any solution of (1.1). The cases \(S = Id\) and \(S \neq Id\) were studied in [16] and [18], respectively. Let \(r^k = b - Ax^k\). It is easy to show that the following relaxation parameter

\[
\lambda_k = \frac{\langle r^k, Mr^k \rangle}{\|A^T M (b - Ax)\|_S^2}
\]

minimizes \(\|x^k - x^*\|\). Simple calculation show that \(\lambda_k \geq 1/\sigma_1^2\). Therefore, to preserve our convergence analysis we suggest following strategy for picking relaxation parameters

\[
\lambda_k = \min \left\{ \frac{\langle r^k, Mr^k \rangle}{\|A^T M (b - Ax)\|_S^2}, \frac{2}{\sigma_1^2} \right\} \text{ for } k = 1, 2, \ldots \tag{3.2}
\]

In Table 2 we demonstrate the effect of using this strategy.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>(C = \mathbb{R}^n)</th>
<th>(C = \mathbb{R}^n_+)</th>
<th>(C = [0,1]^n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landweber</td>
<td>0.2005</td>
<td>0.1642</td>
<td>0.1642</td>
</tr>
<tr>
<td>Cimmino</td>
<td>0.1902</td>
<td>0.1424</td>
<td>0.1424</td>
</tr>
<tr>
<td>CAV</td>
<td>0.1903</td>
<td>0.1425</td>
<td>0.1425</td>
</tr>
<tr>
<td>DROP</td>
<td>0.1902</td>
<td>0.1424</td>
<td>0.1424</td>
</tr>
<tr>
<td>Landweber</td>
<td>0.2488</td>
<td>0.2022</td>
<td>0.2022</td>
</tr>
<tr>
<td>Cimmino</td>
<td>0.2757</td>
<td>0.1975</td>
<td>0.1975</td>
</tr>
<tr>
<td>CAV</td>
<td>0.2756</td>
<td>0.1974</td>
<td>0.1974</td>
</tr>
<tr>
<td>DROP</td>
<td>0.2757</td>
<td>0.1975</td>
<td>0.1975</td>
</tr>
</tbody>
</table>

References