Semi-Convergence and Relaxation Parameters for a Class of SIRT Algorithms

Per Christian Hansen  Technical University of Denmark

Joint work with:
• Tommy Elfving  Linköping University, Sweden
• Touraj Nikazad  Iran University of Science and Technology
• Maria Saxild-Hansen  Technical University of Denmark
Overview of Talk

- Inverse problems and reconstruction algorithms
- Iterative SIRT methods and their semi-convergence
- Strategies for the relaxation parameter (step size)
- A few results
- If time permits: AIR Tools – a new MATLAB® package
Inverse Problems

Goal: use measured data to compute “hidden” information.

Our model: $Ax = b$

Blurring process

Sharp image

Data / blurred image
Tomography = Our Main Application Area

Image reconstruction from projections

Medical scanning

Mapping of metal grains
The Origin of Tomography


Main result:
An object can be perfectly reconstructed from a full set of projections.

NOBELFÖRSAAMLINGEN KAROLINSKA INSTITUTET
THE NOBEL ASSEMBLY AT THE KAROLINSKA INSTITUTE
11 October 1979

The Nobel Assembly of Karolinska Institutet has decided today to award the Nobel Prize in Physiology or Medicine for 1979 jointly to

Allan M Cormack and Godfrey Newbold Hounsfield

for the "development of computer assisted tomography".
Setting Up the Algebraic Model

Damping of $i$-th X-ray through domain:

$$b_i = \int_{ray_i} \chi(s) \, d\ell,$$

where $\chi(s)$ is the attenuation coefficient.

Discretization leads to a large, sparse, ill-conditioned system:

$$A x = b,$$

where $b = b^* + e$

- Geometry
- Projections
- Image
- Noise

$$b^* = A x^*$$
Analogy: the “Sudoku” Problem – 数独

Infinitely many solutions ($c \in \mathbb{R}$):

$$
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
\end{pmatrix}
=
\begin{pmatrix}
3 \\
7 \\
4 \\
6 \\
\end{pmatrix}
$$

Prior: solution is integer and non-negative
Some Large-Scale Inversion Algorithms

Transform-Based Methods
The forward problem is formulated as a certain transform → formulate a stable way to compute the inverse transform. Example: the inverse Radon transform for tomography.

Krylov Subspace Methods
Use the forward model to produce a Krylov subspace → inversion amounts to projecting on this “signal subspace” & using prior information. Examples: CGLS, RRGMRES.

Algebraic Iterative Methods
Formulate the forward problem as a discretized problem → inversion amounts to solving $A x = b$ using algebraic properties of $A$ & using prior information.
Some Algebraic Iterative Methods

**ART – Algebraic Reconstruction Techniques**
- Kaczmarz’s method + variants.
- *Sequential* row-action methods that update the solution using one row of $A$ at a time.

**SIRT – Simultaneous Iterative Reconstruction Techniques**
- Landweber, Cimmino, CAV, DROP, SART, ...
- These methods use all the rows of $A$ *simultaneously* in one iteration (i.e., they are based on matrix multiplications).

**Making the methods useful**
- Relaxation parameter (step length) choice.
- Stopping rules.
- Nonnegativity constraints.
SIRT Methods

The general form:

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

Some methods use the row norms \( \|a^i\|_2 \).

Landweber: \( T = I \) and \( M = I \).

Cimmino: \( T = I \) and \( M = D = \frac{1}{m} \text{diag}\left(\frac{1}{\|a^i\|_2^2}\right) \).

CAV (component averaging method): \( T = I \) and \( M = D_S = \text{diag}\left(\frac{1}{\|a^i\|_S^2}\right) \) with \( S = \text{diag}(\text{nnz}(\text{column } j)) \).

DROP: \( T = S^{-1} \) and \( M = mD \).

SART: \( T = \text{diag}(\text{row sums})^{-1} \) and \( M = \text{diag}(\text{column sums})^{-1} \).
Semi-Convergence of the SIRT Methods

During the first iterations, the iterates $x^k$ capture the “important” information in the noisy right-hand side $b$.

- In this phase, the iterates $x^k$ approach the exact solution.

At later stages, the iterates start to capture undesired noise components.

- Now the iterates $x^k$ diverge from the exact solution and they approach the undesired solution $A^{-1}b$ or $A^\dagger b$.

The iteration number $k$ plays the role of the regularization parameter. This behavior is called *semi-convergence*.

Illustration of Semi-Convergence

\[ x^{(0)} \rightarrow x^{(8)} \rightarrow x^{(20)} \rightarrow x^{(40)} \rightarrow x^{(120)} \rightarrow x^{(250)} \rightarrow A_M^+b \]

\[ x_{\text{exact}} \]
Another Look at Semi-Convergence

Notation: \( b = A x^* + e, \ x^* = \text{exact solution}, \ e = \text{noise}. \)

Initial iterations: the error \( ||x^* - x^k||_2 \) decreases.

Later: the error increases as \( x^k \rightarrow \arg\min_x ||A x - b||_M. \)

The minimum error is independent of both \( \lambda \) and the method.
Analysis of Semi-Convergence

Let \( \bar{x} \) be the solution to the noise-free problem:

\[
\bar{x} = \arg \min_{x \in \mathcal{C}} \frac{1}{2} \| A x - \bar{b} \|_M^2, \quad \bar{b} = \text{pure data}
\]

and let \( x^k \) denote the iterates when applying SIRT to \( \bar{b} \). Then

\[
\| x^k - x^* \|_2 \leq \| x^k - \bar{x}^k \|_2 + \| \bar{x}^k - \bar{x} \|_2.
\]

We need the SVD: \( M^{1/2} A = U \Sigma V^T \) \text{ Assume rank}(A) = n.

The unprojected case is “easy;” \( x^k \) is a filtered SVD solution:

\[
x^k = \sum_{i=1}^n \varphi_i^{[k]} \frac{u_i^T M^{1/2} b}{\sigma_i} v_i, \quad \varphi_i^{[k]} = 1 - (1 - \lambda \sigma_i^2)^k.
\]
The filter factors dampen the “inverted noise” $u_i^T (M^{rac{1}{2}} e)/\sigma_i$.

$\lambda \sigma_i^2 \ll 1 \Rightarrow \varphi_i^{[k]} \approx k \lambda \sigma_i^2 \Rightarrow k$ and $\lambda$ play the same role.
Projected Alg. Noise Error (proof: see paper)

The noise error in projected SIRT is bounded above by

\[ \| x^k - \bar{x}^k \|_2 \leq \frac{\sigma_1 \lambda_0}{\sigma_n \lambda_{k-1}} \Psi^k (\lambda_{k-1}) \| M^{1/2} \delta b \|_2, \]

with

\[ \Psi^k (\lambda) \equiv \frac{1 - (1 - \lambda \sigma_n^2)^k}{\sigma_n}. \]

When \( \lambda_k = \lambda \) for all \( k \) we obtain

\[ \| x^k - \bar{x}^k \|_2 \leq \frac{\sigma_1}{\sigma_n} \Psi^k (\lambda) \| M^{1/2} \delta b \|_2, \]

and as long as \( \lambda \sigma_n^2 \ll 1 \) we have

\[ \| x^k - \bar{x}^k \| \approx \lambda k \sigma_1 \| M^{1/2} \delta b \|_2, \]

showing that \( k \) and \( \lambda \) play the same role for suppressing the noise.
Projected Alg. Iteration Error \((\text{proof: see paper})\)

The iteration error in projected SIRT is bounded above by

\[
\|\bar{x}^k - \bar{x}\|_2 \leq \sigma_n \Phi^k(\lambda_{k-1}) \|x^0 - \bar{x}\|_2,
\]

with

\[
\Phi^k(\lambda) \equiv \frac{(1 - \lambda \sigma_n^2)^k}{\sigma_n}.
\]

Our bound have pessimistic factors, but track well the actual errors:

- **NE**: actual noise error
- **NE-b**: our bound without the factor $\sigma_1/\sigma_n$
- **IE**: actual iteration error
- **IE-b**: our bound without the factor $\|x^0 - \bar{x}\|_2$
Choosing the SIRT Relaxation Parameter

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

Goal: fast semi-convergence to the minimum error.

**Training.** Using a noisy test problem, find the fixed \( \lambda_k = \lambda \) that gives fastest semi-convergence to the minimum error.

**Line search** (Dos Santos, Appleby & Smolarski, Dax). Minimize the error \( \|x^k - x^*\|_2 \) in each iteration – must assume that \( A x = b \) is consistent. When \( T = I \) we get:

\[ \lambda_k = (r^k)^T M r^k / \| A^T M r^k \|_2^2, \quad r^k = b - A x^k. \]
Preparation for More Insight ...

The function (which appears in the analysis)

\[ g_{k-1}(y) = (2k - 1)y^{k-1} - (y^{k-2} + \ldots + y + 1) \]

has a unique real root \( \zeta_k \in (0, 1) \). The roots satisfy

\[ 0 < \zeta_k < \zeta_{k+1} < 1 \quad \text{and} \quad \lim_{k \to \infty} = 1 \]

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \zeta_k )</th>
<th>( k )</th>
<th>( \zeta_k )</th>
<th>( k )</th>
<th>( \zeta_k )</th>
<th>( k )</th>
<th>( \zeta_k )</th>
<th>( k )</th>
<th>( \zeta_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.3333</td>
<td>7</td>
<td>0.8156</td>
<td>12</td>
<td>0.8936</td>
<td>17</td>
<td>0.9252</td>
<td>22</td>
<td>0.9424</td>
</tr>
<tr>
<td>3</td>
<td>0.5583</td>
<td>8</td>
<td>0.8392</td>
<td>13</td>
<td>0.9019</td>
<td>18</td>
<td>0.9294</td>
<td>23</td>
<td>0.9449</td>
</tr>
<tr>
<td>4</td>
<td>0.6719</td>
<td>9</td>
<td>0.8574</td>
<td>14</td>
<td>0.9090</td>
<td>19</td>
<td>0.9332</td>
<td>24</td>
<td>0.9472</td>
</tr>
<tr>
<td>5</td>
<td>0.7394</td>
<td>10</td>
<td>0.8719</td>
<td>15</td>
<td>0.9151</td>
<td>20</td>
<td>0.9366</td>
<td>25</td>
<td>0.9493</td>
</tr>
<tr>
<td>6</td>
<td>0.7840</td>
<td>11</td>
<td>0.8837</td>
<td>16</td>
<td>0.9205</td>
<td>21</td>
<td>0.9396</td>
<td>26</td>
<td>0.9513</td>
</tr>
</tbody>
</table>
Parameter-Choice: Limit the Noise Error

Assume that $0 < \lambda_{i-1} \leq \lambda_i$ in steps $1, \ldots, k - 1$; then

$$
\|x^k - \bar{x}^k\|_2 \leq \frac{\sigma_1 \lambda_0}{\sigma_n \sqrt{\lambda_{k-1}}} \frac{1 - \zeta_k^k}{\sqrt{1 - \zeta_k}} \|M^{1/2} \delta b\|_2
$$

**Strategy $\Psi_1$:** choose $\lambda_0 = \lambda_1 = \sqrt{2}/\sigma_1^2$ and

$$
\lambda_k = \frac{2}{\sigma_1^2} (1 - \zeta_k), \quad k = 2, 3, \ldots
$$

**Strategy $\Psi_2$:** choose $\lambda_0 = \lambda_1 = \sqrt{2}/\sigma_1^2$ and

$$
\lambda_k = \frac{2}{\sigma_1^2} \frac{1 - \zeta_k}{(1 - \zeta_k^k)^2}, \quad k = 2, 3, \ldots
$$

Both are diminishing: $\lambda_k \to 0$ such that $\sum_k \lambda_k = \infty$. 
Our New Strategies: What we Achieve

As a result:

\[ \|x^k - \bar{x}_k\|_2 \leq \frac{\sigma_1^2 \lambda_0}{\sigma_n \sqrt{2}} \frac{1 - \zeta_k^k}{1 - \zeta_k} \|M^{1/2} \delta b\|_2 \quad \text{for strategy } \Psi_1 \]

\[ \|x^k - \bar{x}_k\|_2 \leq \frac{\sigma_1^2 \lambda_0}{\sigma_n \sqrt{2}} \frac{(1 - \zeta_k^k)^2}{1 - \zeta_k} \|M^{1/2} \delta b\|_2 \quad \text{for strategy } \Psi_2 \]
Error Histories for Cimmino Example

All three strategies give fast semi-convergence:
- The fixed $\lambda$ requires training and thus a realistic test problem.
- The Dos Santos line search often gives a ‘zig-zag’ behavior.
- Our new strategy clearly controls the noise propagation.
Numerical Results (SNARK model problem)

P-Cimmino, $\eta = 0.01$

- Opt. $\lambda$
- $\Psi_1$
- $\Psi_2$

Relative error vs. Iteration number

P-DROP, $\eta = 0.01$

- Opt. $\lambda$
- $\Psi_1$
- $\Psi_2$

Relative error vs. Iteration number

P-Cimmino, $\eta = 0.05$

P-Cimmino, $\eta = 0.08$

P-DROP, $\eta = 0.05$

P-DROP, $\eta = 0.08$
Conclusions

- We have verified the observed simiconvergence of the standard and the projected SIRT methods.
- We proposed two new strategies for choosing $\lambda_k$.
- Our strategies control the noise component of the error.
- In case of noise-free data our strategies give convergence to the problem $\min_{x \in C} \| A x - b \|_M^2$.
- Our strategies also work for consistent and inconsistent systems, for rank-deficient matrices, and SIRT methods with $T \neq I$.
- They are implemented in the MATLAB package AIR Tools.