Semi-Convergence Properties of Kaczmars' Algorithm

Per Christian Hansen Technical University of Denmark

Joint work with:

- Tommy Elfving
  Linköping University, Sweden
- Touraj Nikazad
  Iran University of Science and Technology

HDtomo
HIGH - DEFINITION TOMOGRAPHY

DTU Compute
Department of Applied Mathematics and Computer Science

Overview of Talk

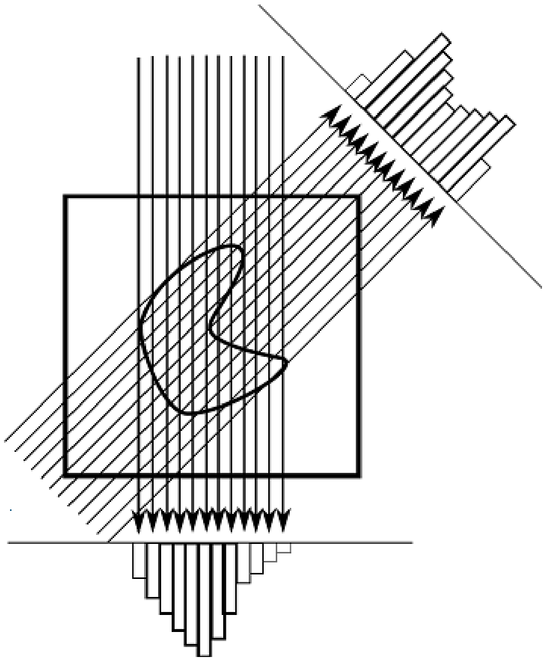
- Tomography $\rightarrow$ $A\,x = b$
- Algebraic iterative methods
- Semi-convergence
- Analysis and results

Why consider such a simple method?

- Suited for very-large-scale problems.

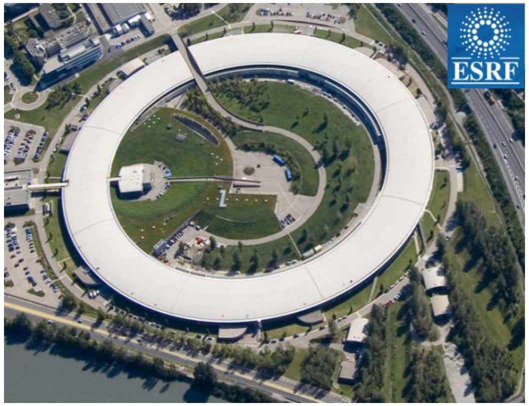
Tomography = Our Main Application Area

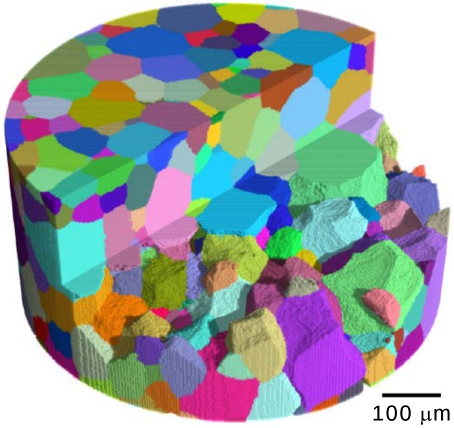
Image reconstruction from projections



Medical scanning

Mapping of metal grains

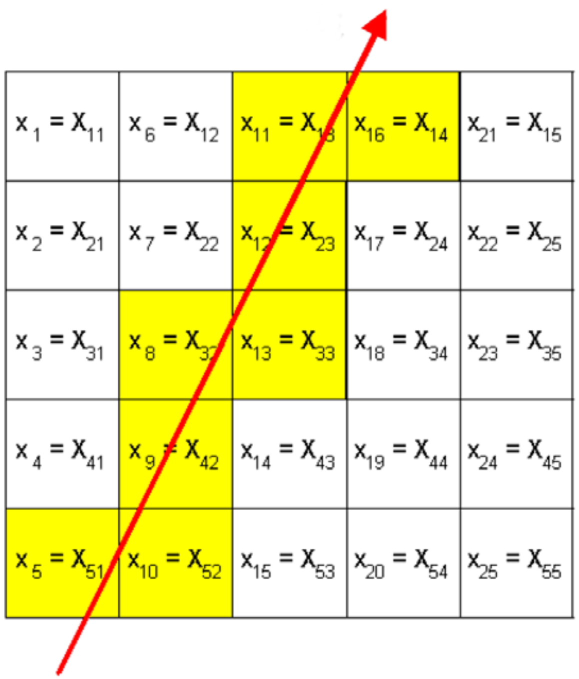




Setting Up the Algebraic Model

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

$$b_i = \int_{\mathrm{ray}_i} \chi(\mathbf{s})\, d\ell, \quad \chi(\mathbf{s}) = \text{attenuation coef.}$$



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

$$A\,x = b \qquad b = \bar{b} + e$$

Geometry (A), Image (x), Projections (b), Noise (e)

$$\bar{b} = A\,\bar{x}$$

Some Large-Scale Reconstruction Algorithms

Bayesian Methods

My knowledge here is very limited ...

Transform-Based Methods

The forward problem is formulated as a certain transform
$\rightarrow$ find a stable way to compute the inverse transform.

Examples: the inverse Radon transform for tomography
$\rightarrow$ filtered back-projection, FDK.

Algebraic Iterative Methods

The forward problem is formulated as a discretized problem
$\rightarrow$ solve $A\,x = b$ iteratively using prior information.

Examples: Cimmino, Kaczmarz, CGLS.

This work

Classical Algebraic Iterative Methods

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):

$$x \leftarrow \mathcal{P}\big(x + \omega\, A^T M (b - A\,x)\big)$$

$$\mathcal{P} = \text{projection on a convex set (e.g., } x \geq 0)$$

ART – Algebraic Reconstruction Techniques

- Kaczmarz's method + variants.
- *Sequential* row-action methods that update the solution using one row of *A* at a time:

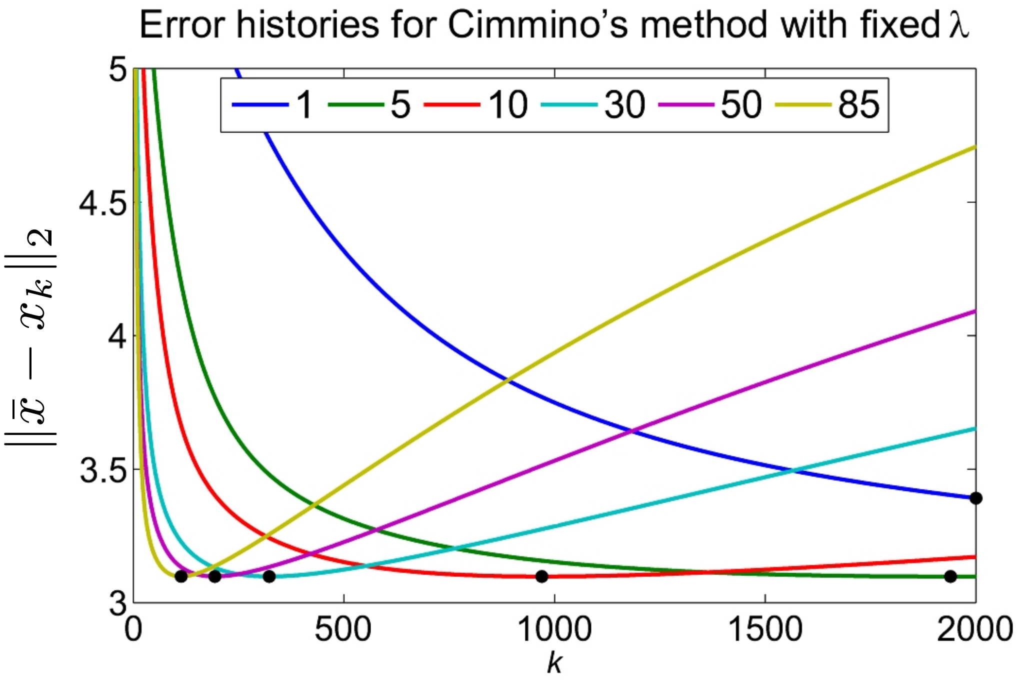
$$x \leftarrow \mathcal{P}\left(x + \omega \frac{b_i - a_i^T x}{\|a_i\|_2^2} a_i\right) \qquad a_i^T = i\text{th row of } A$$

Semi-Convergence

Notation: $b = A\,\bar{x} + e$, $\bar{x}$ = exact solution, e = noise.

Initial iterations: the error $\|\bar{x} - x_k\|_2$ decreases.

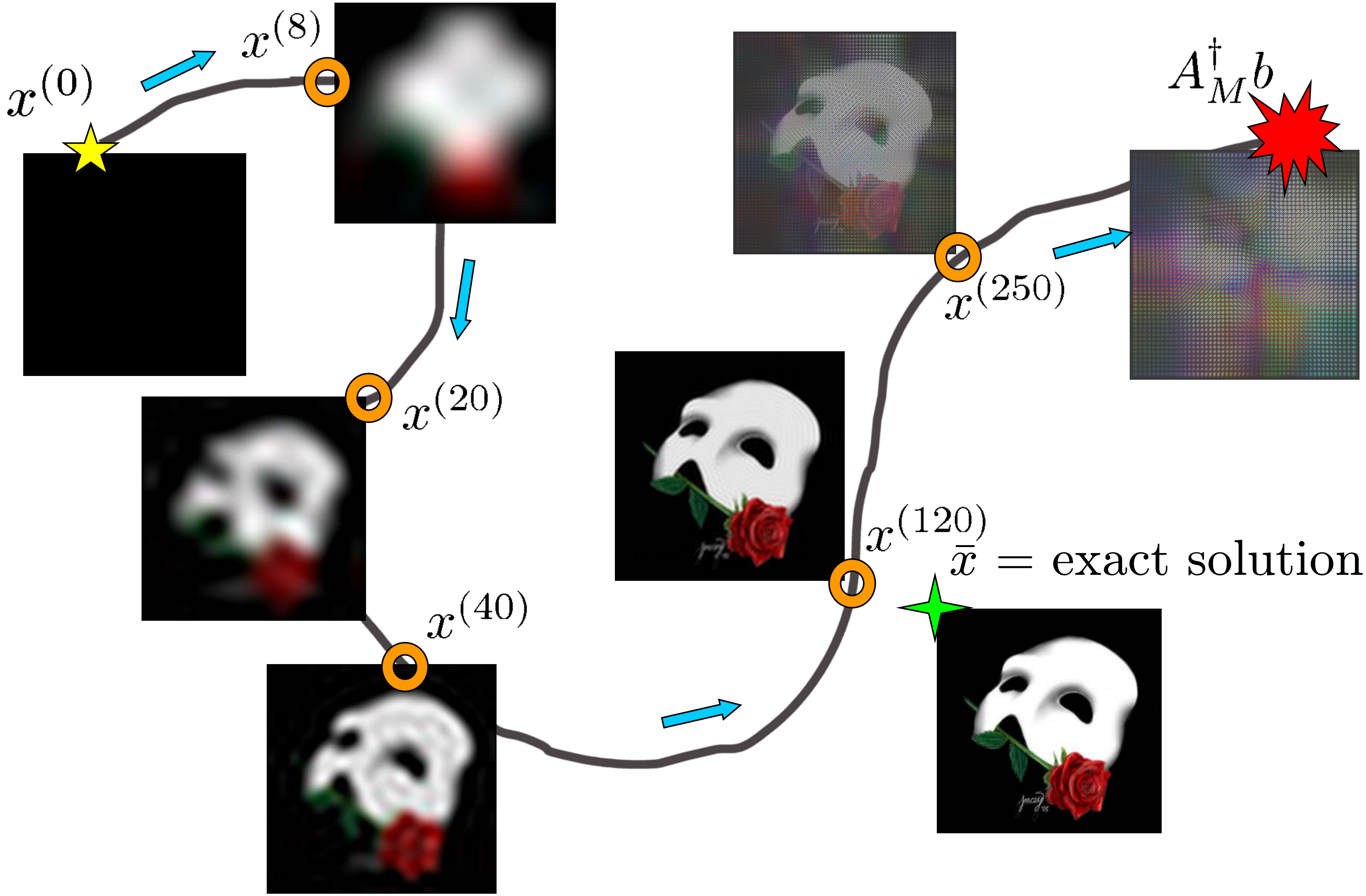
Later: the error increases as $x_k \to$ (weighted) least squares solution.



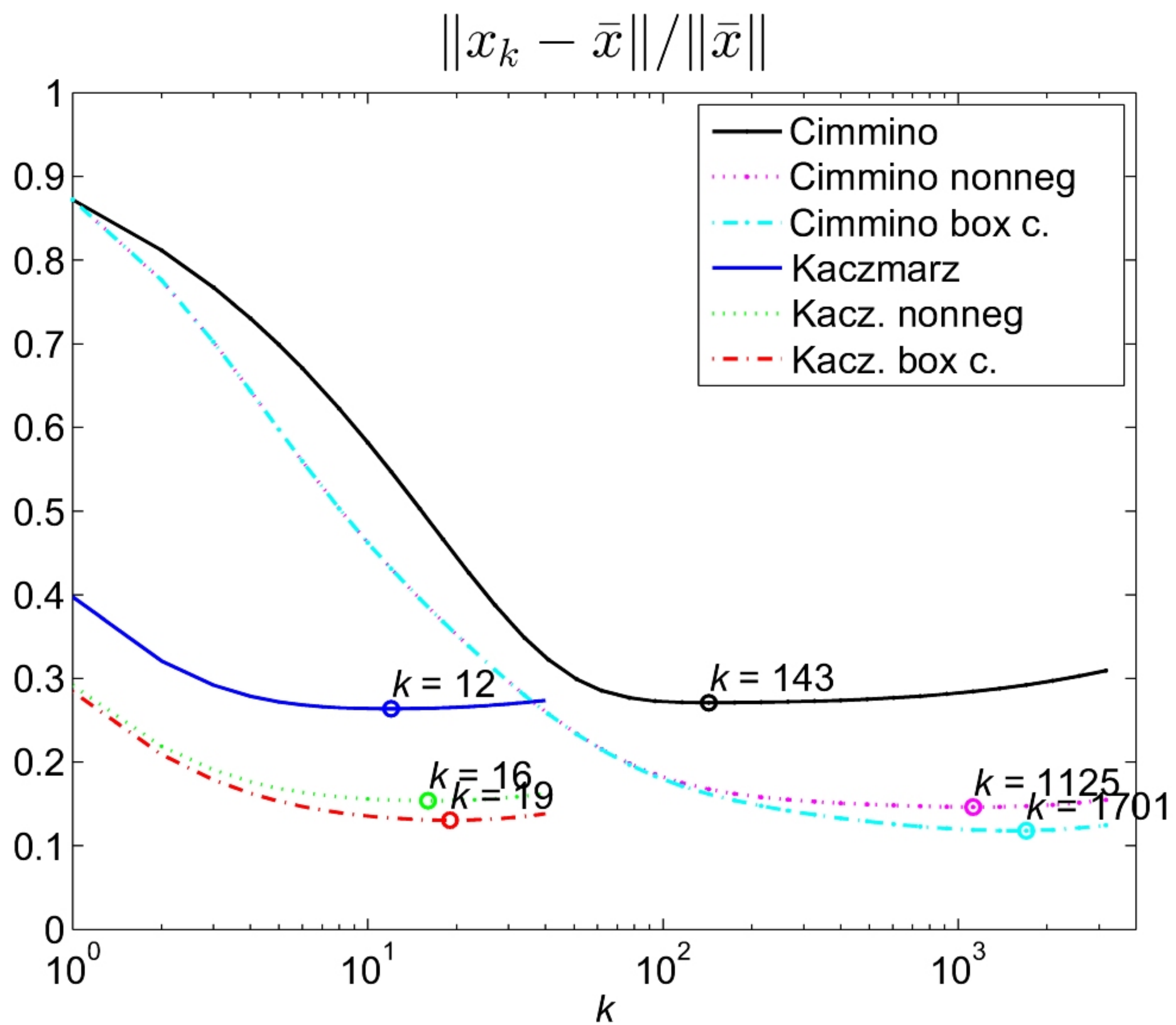
A few references:

- F. Natterer, *The Mathematics of Computerized Tomography* (1986)
- A. van der Sluis & H. van der Vorst, *SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems* (1990)
- M. Bertero & P. Boccacci, *Inverse Problems in Imaging* (1998)
- M. Kilmer & G. W. Stewart, *Iterative Regularization And Minres* (1999)
- H. W. Engl, M. Hanke & A. Neubauer, *Regularization of Inverse Problems* (2000)

Illustration of Semi-Convergence



Semi-Convergence of SIRT and ART



- SIRT's semi-convergence is "easy" to show using SVD.
- ART also has semi-convergence; not rigorously proved.
- ART converges much *faster* than SIRT.

<u>Example</u>

parallel-beam with 200×200 phantom and 60 projections

Analysis of Semi-Convergence for ART

Let $\bar{x}$ be the solution to the noise-free problem, and let $\bar{x}^k$ denote the iterates when applying ART to $\bar{b}$. Then

$$\|x_k - \bar{x}\|_2 \;\square\; \|x_k - \bar{x}_k\|_2 + \|\bar{x}_k - \bar{x}\|_2 \ .$$

Noise error

Iteration error

The convergence theory for ART is well established and ensures that the iteration error $\bar{x}_k - \bar{x}$ goes to zero.

Our concern here is the noise error $e_k^{\mathrm{N}} = x_k - \bar{x}_k$. We wish to establish that it increases, and how fast.

A Word on the Iteration Error

Strohmer & Vershynin (2009): Known estimates of convergence rates are based on quantities of A that are hard to compute and difficult to compare with convergence estimates of other iterative methods.

What numerical analysts would like to have is estimates of the convergence rate with respect to standard quantities such as $||A||$ and $||A^{-1}||$.

The difficulty: the rate of convergence for ART depends on the ordering of the equations, while $||A||$ and $||A^{-1}||$ are independent of the ordering.

With *random* selection of the rows, the expected behavior is:

$$\left(1 - \frac{2k}{\mathrm{cond}(A)^2}\right) \|\bar{x}_0 - \bar{x}\|_2^2 \;\square$$

$$\mathcal{E}\left(\|\bar{x}^k - \bar{x}\|_2^2\right) \;\square\; \left(1 - \frac{1}{\mathrm{cond}(A)^2}\right)^k \|\bar{x}_0 - \bar{x}\|_2^2.$$

Note: $AA^T =$ diagonal matrix $\Rightarrow$ convergence in one sweep!

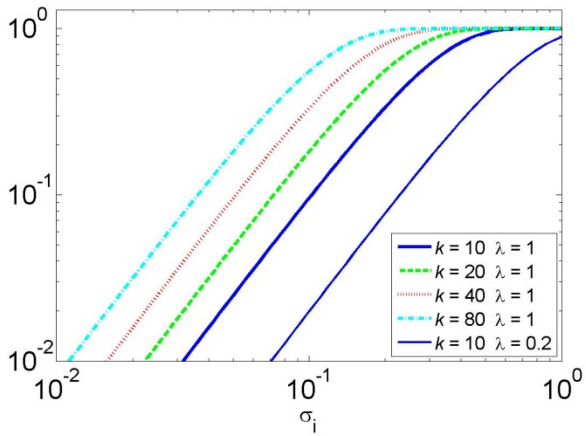
Sidetrack: Noise Error for SIRT

The unprojected case:
x_k is a filtered SVD solution:

$$x_k \quad = \quad \sum_{i=1}^{n} \varphi_i^{[k]} \, \frac{u_i^T M^{1/2} b}{\sigma_i} \, v_i$$

$$\varphi_i^{[k]} \quad = \quad 1 - \left(1 - \omega\, \sigma_i^2\right)^k .$$

Filter factors $\varphi_i^{[k]} = 1 - \left(1 - \omega\,\sigma_i^2\right)^k$



With projection an SVD analysis is not possible; we obtain:

$$\|x^k - \bar{x}^k\|_2 \;\square\; \frac{\sigma_1}{\sigma_n} \, \frac{(1 - \omega\sigma_n^2)^k}{\sigma_n} \, \|M^{1/2}\delta b\|_2$$

and for $\omega\,\sigma_n^2 \ll 1$ we have:

$$\|x^k - \bar{x}^k\|_2 \approx \omega\, k\, \sigma_1 \|M^{1/2}\delta b\|_2.$$

Elfving, H,
Nikazad, 2012

Noise Error for ART – No Projection

Recall: ART is equivalent to applying SOR to $A\,A^T y = b$, $x = A^T y$. We introduce the splitting:

$$AA^T = L + D + L^T, \qquad M = (D + \omega L)^{-1},$$

where L is strictly lower triangular and $D = \mathrm{diag}(\|a_i\|_2^2)$. Then:

$$x_{k+1} = x_k + \omega A^T M\,(b - A\,x_k)\ .$$

We also introduce

$$e = b - \bar{b} = \text{noise in data}, \qquad Q = I - \omega A^T M\,A\ .$$

Then simple manipulations show that the noise error is given by

$$e_k^{\mathrm{N}} = x_k - \bar{x}_k = Q\ e_{k-1}^{\mathrm{N}} + \omega A^T M\,e = \omega \sum_{j=1}^{k-1} Q^j A^T M\,e\ .$$

Noise Error Analysis - I

Let P = projection matrix on range(A^T) and $u = A^T M e$; then:

$$\begin{aligned} Q^k u &= Q^k P u = (I - \omega B)(I - \omega B) \cdots (I - \omega B) P u \\ &= (I - \omega B) P (I - \omega B) P \cdots (I - \omega B) P u \;=\; (QP)^k u. \end{aligned}$$

Hence

$$e_k^{\mathrm{N}} = \omega \sum_{j=0}^{k-1} Q^j A^T M e = \omega \sum_{j=0}^{k-1} (QP)^j A^T M e$$

and, with $q = \|QP\|_2$ and $\delta = \|A^T M\, e\|_2$,

$$\|e_k^{\mathrm{N}}\|_2 \le \omega\,\delta \left\| \textstyle\sum_{j=0}^{k-1} (QP)^j \right\|_2 \le \omega\,\delta \sum_{j=0}^{k-1} q^j = \omega\,\delta \frac{1 - q^k}{1 - q}.$$

Noise Error Analysis - II

Lemma

$$q^2 = 1 - \bar{\omega}\sigma_r^2, \quad \bar{\omega} = \omega(2-\omega), \quad \sigma_r = \text{smallest nonzero s.v. of } D^{1/2}MA.$$

Taylor

$$\begin{aligned} q &= \sqrt{1-\bar{\omega}\sigma_r^2} = 1 - \tfrac{1}{2}\bar{\omega}\sigma_r^2 + O(\sigma_r^4) \\ \frac{1-q^k}{1-q} &= \frac{1-(1-\frac{1}{2}\bar{\omega}\sigma_r^2 + O(\sigma_r^4))^k}{\frac{1}{2}\bar{\omega}\sigma_r^2 + O(\sigma_r^4))} \\ &= \frac{1-(1-k\,\frac{1}{2}\bar{\omega}\sigma_r^2 + O(\sigma_r^4))}{\frac{1}{2}\bar{\omega}\sigma_r^2 + O(\sigma_r^4))} = k + O(\sigma_r^2). \end{aligned}$$

These results lead to the bound

$$\|e_k^{\mathrm{N}}\|_2 \;\square\; \omega\delta\,\frac{1-q^k}{1-q} = \omega\,\delta\,k + O(\sigma_r^2).$$

Noise Error Analysis – A Tighter Bound

Further analysis (see the paper) shows that the noise error in ART is bounded above as:

$$\|e_k^{\mathrm{N}}\|_2 \Box \frac{\delta}{\sigma_r}\,\Psi_k + \mathcal{O}(\sigma_r^2), \qquad \Psi_k = \frac{1-(1-\omega\sigma_r^2)^k}{\sigma_r}.$$

As long as $\omega\sigma_r^2 < 1$ we have

$$\Psi_k \Box \sqrt{\omega}\sqrt{k}$$

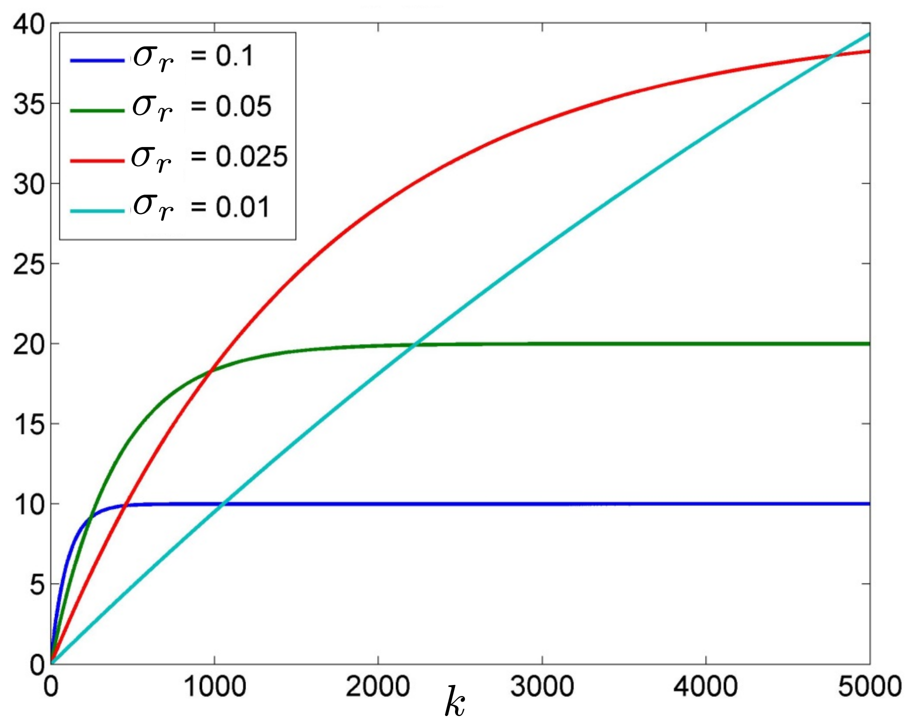
and thus

$$\|e_k^{\mathrm{N}}\|_2 \Box \frac{\sqrt{\omega}\,\delta}{\sigma_r}\sqrt{k} + \mathcal{O}(\sigma_r^2).$$

This also holds for *projected* ART provided that A and P satisfy

$$y \in \mathcal{R}(A^T) \Rightarrow \mathcal{P}(y) \in \mathcal{R}(A^T).$$

Ψ_k for $\omega = 1$



Numerical Results ('paralleltomo' from AIR Tools)

The point of **semi-convergence** arises when noise error ≈ iteration error.

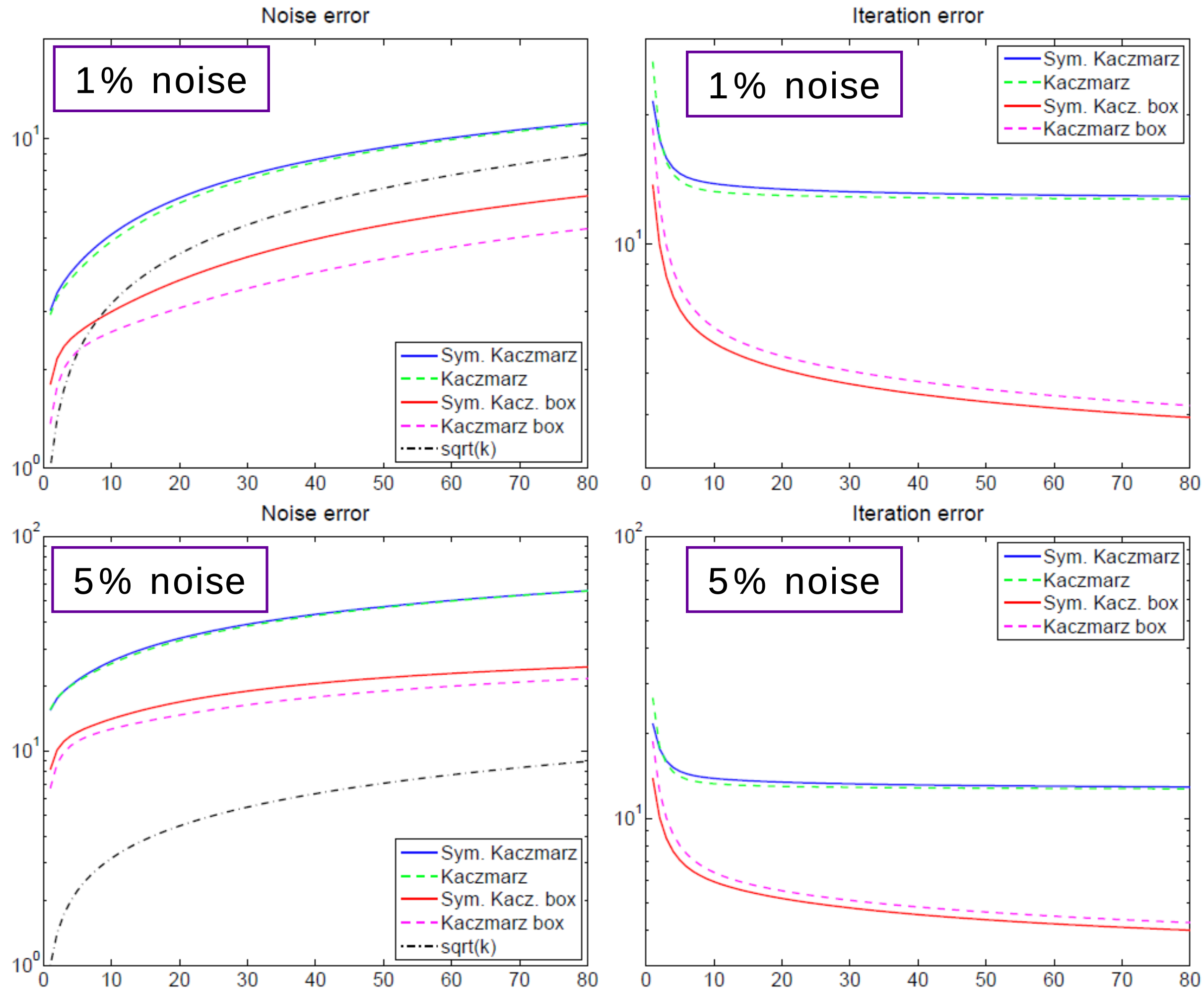
Test problem:

- 200×200 phantom,
- 60 projections at
- 3°,6°,9°,…,180°,
- m = 15,232,
- n = 40,000.

We estimate

$$\sqrt{\omega}\delta/\sigma_r \approx 10^7.$$

Hence our bound is a wild over-estimate but it correctly *tracks* the noise error.



More Insight: SVD Analysis

We consider two specific SIRT and ART algorithms.

Cimmino is an unprojected SIRT method:

$$x \leftarrow x + \lambda\, A^T M_{\mathrm{C}} (b - A\, x), \qquad M_{\mathrm{C}} = \mathrm{diag}(\|a_i\|_2^{-2}) \ .$$

Symmetric ART is an unprojected ART method

$$x \leftarrow x + \omega \frac{b_i - a_i^T x}{\|a_i\|_2^2} a_i$$

with the specific row ordering $i = 1, 2, 3, \ldots, n, n{-}1, n{-}2, ..., 1, 2, 3, \ldots$ which can be expressen in "SIRT form" with

$$M_{\mathrm{S}} = (2 - \omega)\,(D + \omega L^T)^{-1} D\,(D + \omega L)^{-1} \ .$$

We can perform an SVD analysis of $M^{1/2} A$ for both methods.

SVD Analysis – How To

We need this SVD:

$$M^{1/2}A = U\,\Sigma\,V^T \ .$$

Then

$$x_k = \sum_{i=1}^{n} \phi_i^{(k)} \, \frac{u_i^T v}{\sigma_i} \, v_i \ ,$$

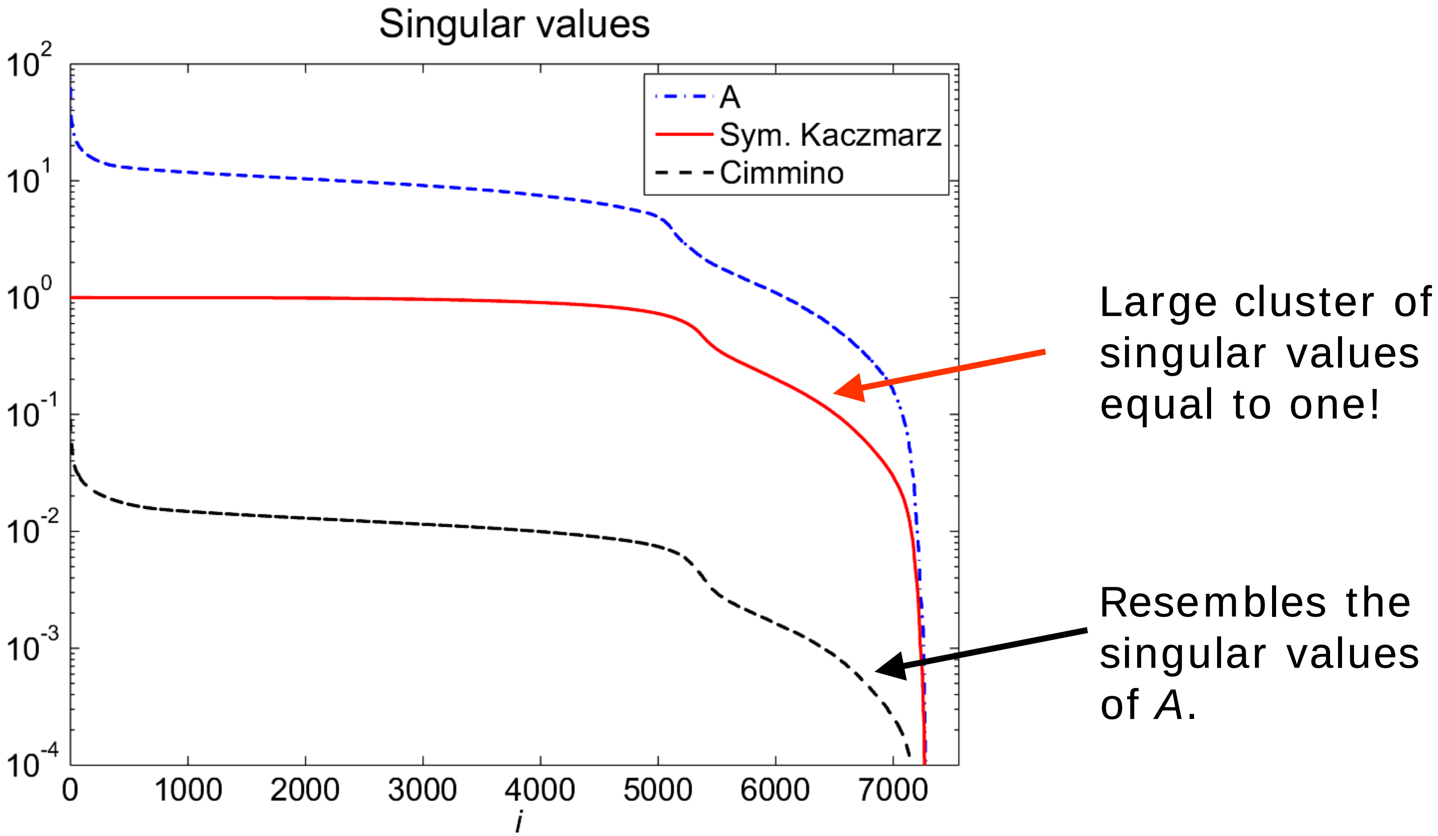
with the filter factors

$$\phi_i^{(k)} = 1 - (1 - \omega\sigma_i^2)^k, \qquad i = 1, 2, \ldots, n \ .$$

The iterates correspond to "spectral filtering."

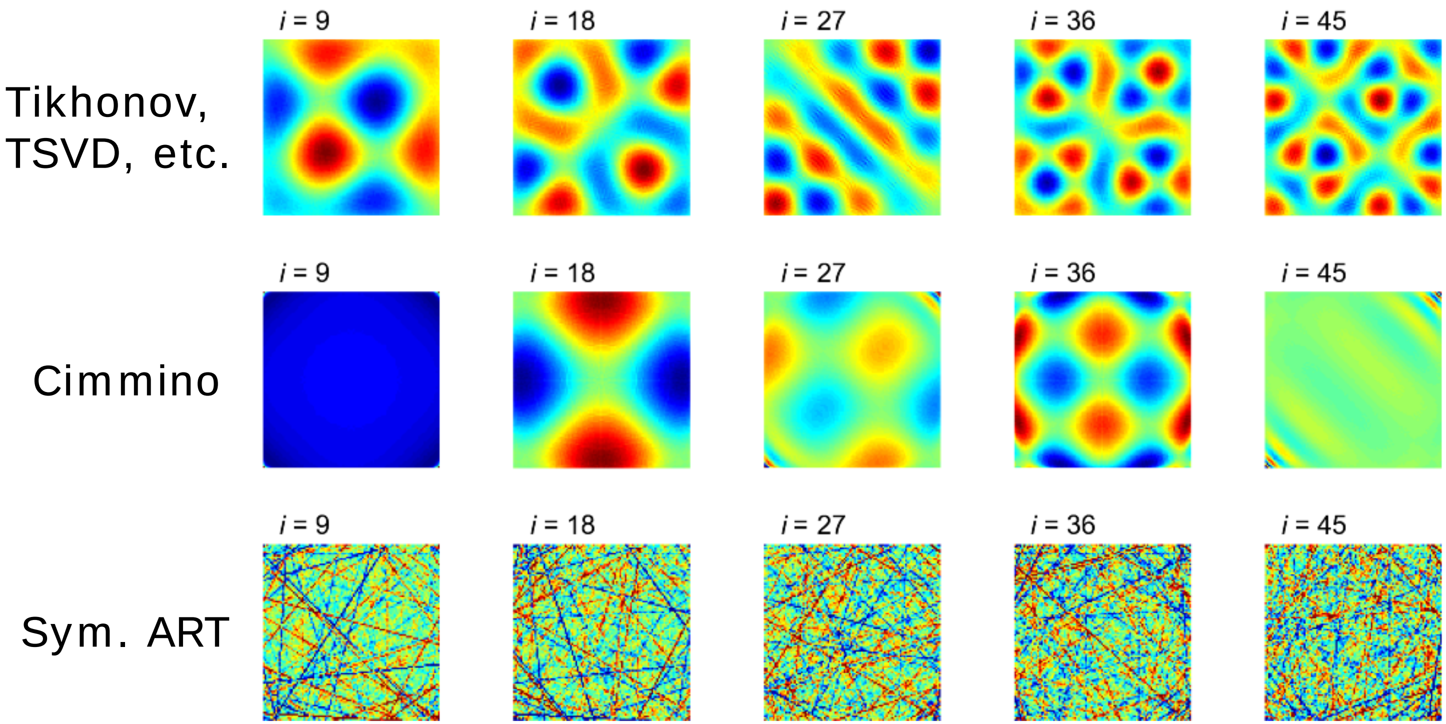
Singular Values

The singular values of A and those of $M_{\mathrm{C}}^{1/2} A$ and $M_{\mathrm{S}}^{1/2} A$ associated with Cimmino and Symmetric ART ($\lambda = 1$):



Singular Vectors

Some singular vectors of A, $M_{\mathrm{C}}^{1/2}A$ and $M_{\mathrm{S}}^{1/2}A$, shown as 2D images:



Cimmino gives ”smooth” solutions, similar to Tikhonov and Truncated SVD.
Symmetric ART can give solutions with fine-grained structure.

Conclusions

- Semi-convergence is well established for SIRT and CGLS.
- We provide a first step toward ditto for ART:
  - Analysis of the convergence of the noise error – we give an upper bound for the noise error (lower bound = ???).
  - Insight into structure of the singular values and vectors.
- More details + block methods: see our paper.
- Next steps: more insight, choice of relaxation parameter ω.