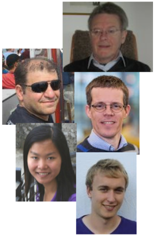
Convergence Stories of Algebraic Iterative Reconstruction

Per Christian Hansen
DTU Compute, Technical University of Denmark



Joint work with

Tommy Elfving – Linköping University
Touraj Nikazad – Iran Univ. Scence & Technology
Michiel E. Hochstenbach – TU Eindhoven
Yiqiu Dong & Nicolai A. B. Riis – DTU Compute

Software with Algebraic Iterative Methods

- **ASTRA Toolbox:** a MATLAB & Python toolbox of GPU primitives for 2D and 3D tomography, developed at University of Antwerp and CWI, Amsterdam. https://www.astra-toolbox.com/
- **SNARK14:** a C++ system for reconstruction of 2D images from 1D projections, developed at City Univ. New York and Univ. Nacional Autónoma de México. https://turing.iimas.unam.mx/SNARK14M/
- **jSNARK:** a C++ programming system for the reconstruction of 2D and 3D images from their projections, developed at City University of New York. http://jsnark.sourceforge.net/
- **TIGRE:** Matlab and Python libraries for tomographic iterative GPU-based reconstruction, developed at the University of Bath and CERN. https://github.com/CERN/TIGRE/
- **AIR Tools II:** a Matlab toolbox of algebraic iterative reconstruction methods, developed at the Technical Univ. of Denmark and the Univ. of Manchester. http://people.compute.dtu.dk/pcha/AIRtoolsII/
- **FAIR Tools:** a port to Fortran 90 of parts of the AIR Tools II package, developed at the Technical University of Denmark. https://github.com/BartvLith/fortran_AIRtools/

Goal of my Talk

Algebraic iterative reconstruction methods (Kaczmarz, Cimmino, etc.) are successfully used in *computed tomography:*

- Very flexible – no assumptions about the CT scanning geometry.
- Easy to incorporate convex constraints (e.g., nonneg./box constraints).

Goal of my Talk

Algebraic iterative reconstruction methods (Kaczmarz, Cimmino, etc.) are successfully used in *computed tomography:*

- Very flexible – no assumptions about the CT scanning geometry.
- Easy to incorporate convex constraints (e.g., nonneg./box constraints).

Both of these statements are true:

- We know a lot about the convergence – for exact data.
- We know so little about the convergence – for noisy data.

Goal of my Talk

Algebraic iterative reconstruction methods (Kaczmarz, Cimmino, etc.) are successfully used in *computed tomography:*

- Very flexible – no assumptions about the CT scanning geometry.
- Easy to incorporate convex constraints (e.g., nonneg./box constraints).

Both of these statements are true:

- We know a lot about the convergence – for exact data.
- We know so little about the convergence – for noisy data.

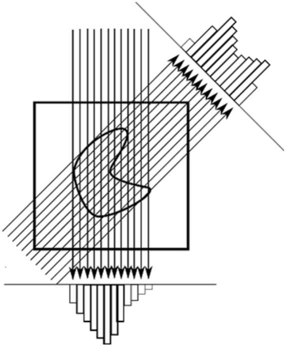
$\boxed{\text{This talk tells the tale of the evolution of convergence theory.}}$

All proofs: see the papers.

X-Ray Tomography and the Radon Transform

The Principle

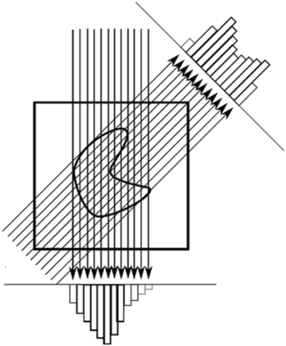
Send X-rays through the object at different angles, and measure the attenuation.

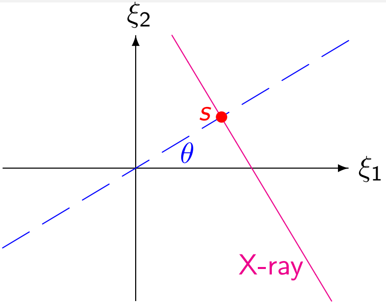


X-Ray Tomography and the Radon Transform

The Principle

Send X-rays through the object at different angles, and measure the attenuation.





$$f(\boldsymbol{\xi}) = \text{2D object/image}, \quad \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$$

$$g(\theta, s) = \mathcal{R}\, f = \text{Radon transform of } f$$

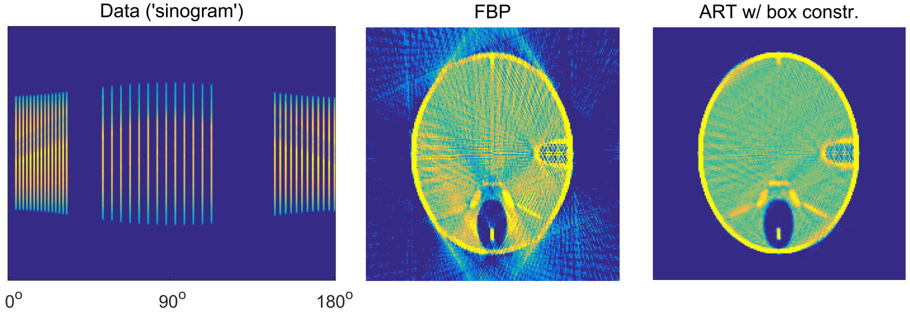
$$= \int_{-\infty}^{\infty} f\left(s \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + \tau \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \right) d\tau$$

$$\mathcal{R}^{-1} = \text{Filtered Back Projection (FBP)}$$

Filtered Back Projection Versus Algebraic Reconstruction

- FBP: fast, low memory, good results with sufficiently many good data.
- But *artifacts* appear with noisy and/or limited data.
- Difficult to incorporate constraints (e.g., nonnegativity).
- Algebraic iterative reconstruction methods are more flexible and adaptive – but require more computational work.

Example with 3% noise and an incomplete set of projection angles:



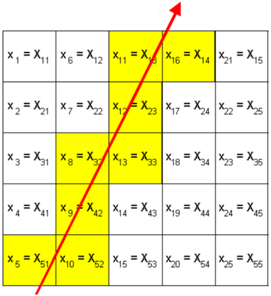
Setting Up the Algebraic Model

Assume each detector pixel is hit by a single X-ray. The Lambert-Beer law says that the damping of the ith X-ray through the domain is a line integral:

$$b_i = \int_{\mathrm{ray}_i} f(\xi)\, d\ell, \qquad f(\xi) = \text{attenuation coef.}$$

Assume $f(\xi)$ is a constant x_j in pixel j, leading to:

$$b_i = \sum a_{ij}\, x_j, \qquad a_{ij} = \begin{cases} \text{length of ray } i \text{ in pixel } j \\ 0 \text{ otherwise.} \end{cases}$$



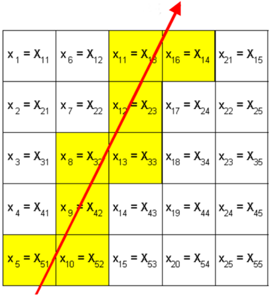
Setting Up the Algebraic Model

Assume each detector pixel is hit by a single X-ray. The Lambert-Beer law says that the damping of the ith X-ray through the domain is a line integral:

$$b_i = \int_{\mathrm{ray}_i} f(\xi)\, d\ell, \qquad f(\xi) = \text{attenuation coef.}$$

Assume $f(\xi)$ is a constant x_j in pixel j, leading to:

$$b_i = \sum a_{ij}\, x_j, \qquad a_{ij} = \begin{cases} \text{length of ray } i \text{ in pixel } j \\ 0 \text{ otherwise.} \end{cases}$$



This leads to a linear system of equations:

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

$A \sim$ measurement geometry,
$x \sim$ reconstruction,
$b \sim$ data.

Note: A is sparse; often we do not store it.

Algebraic Systems and Iterative Methods

Our notation:

$$A\,x = b, \qquad A \in \mathbb{R}^{m\times n}, \qquad x \in \mathbb{R}^{m}, \qquad b \in \mathbb{R}^{m}\ .$$

Algebraic Systems and Iterative Methods

Our notation:

$$A\,x = b, \qquad A \in \mathbb{R}^{m\times n}, \qquad x \in \mathbb{R}^m, \qquad b \in \mathbb{R}^m \ .$$

Two types of algebraic iterative methods:

- **Simultaneous iterations** such as Cimmino's method

$$x^{k+1} = x^k + \omega\, A^T M\,(b - A\,x^k), \quad M = \mathrm{diag}(\|A(i,:)\|_2^{-2}).$$

- **Row-action methods** such as Kaczmarz' method

$$x^{k+1} = x^k + \omega\,\frac{b_i - A(i,:)\,x^k}{\|A(i,:)\|_2^2}\,A(i.;)^T, \qquad i = k \bmod m.$$

Algebraic Systems and Iterative Methods

Our notation:

$$A\,x = b, \qquad A \in \mathbb{R}^{m\times n}, \qquad x \in \mathbb{R}^m, \qquad b \in \mathbb{R}^m \ .$$

Two types of algebraic iterative methods:

- **Simultaneous iterations** such as Cimmino's method

$$x^{k+1} = x^k + \omega\, A^T M\,(b - A\,x^k), \quad M = \mathrm{diag}(\|A(i,:)\|_2^{-2}).$$

- **Row-action methods** such as Kaczmarz' method

$$x^{k+1} = x^k + \omega\, \frac{b_i - A(i,:)\,x^k}{\|A(i,:)\|_2^2}\, A(i.;)^T, \qquad i = k \bmod m.$$

Multiplication with A $\leftrightsquigarrow$ (forward) projector.
Multiplication with A^T $\leftrightsquigarrow$ backprojector.

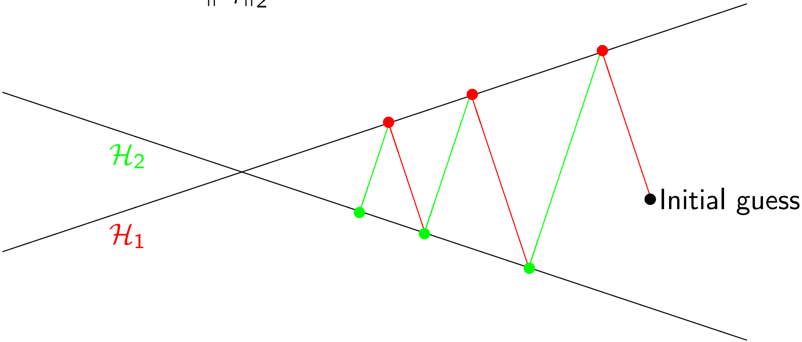
NB: the *implementation* of the backprojector may differ from A^T.

ART: Algebraic Reconstruction Technique = Kaczmarz

Kaczmarz (1937): $x \leftarrow \mathcal{P}_i\, x$ = orthogonal projection on the hyperplane $\mathcal{H}_i$ defined by the ith row a_i^T of A and the corresp. element b_i of the rhs.

Repeat accessing the rows *sequentially*, e.g., in a cyclic fashion:

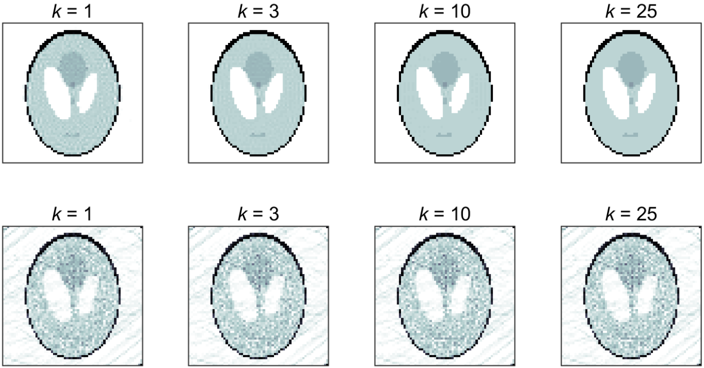
$$x \leftarrow \mathcal{P}_i\, x = x + \frac{b_i - a_i^T x}{\|a_i\|_2^2}\, a_i \ , \qquad i = 1, 2, \ldots, m, 1, 2, \ldots, m, 1, 2, \ldots$$



Can also access the rows in a randomized fashion.

Example of ART Performance

Image size 64×64. Data: 360 projection angles in $[0^\circ, 360^\circ]$, 90 detector pixels (90 rays per projection).



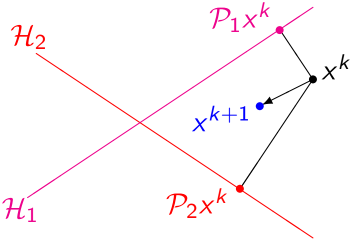
Top: no noise. Bottom: 10% Gaussian noise.

From Sequential to Simultaneous Updates

Cimmino (1938): access all rows *simultaneously* and compute next iterate as the average of the all the projections of the previous iterates:

$$
\begin{aligned}
x^{k+1} &= \frac{1}{m}\sum_{i=1}^{m} \mathcal{P}_i x^k = \frac{1}{m}\sum_{i=1}^{m}\left(x^k + \frac{b_i - a_i^T x^k}{\|a_i\|_2^2}\, a_i\right) \\
&= x^k + \frac{1}{m}\sum_{i=1}^{m} \frac{b_i - a_i^T x^k}{\|a_i\|_2^2}\, a_i = x^k + A^T M\left(b - A\, x^k\right),
\end{aligned}
$$

where we introduced the diagonal matrix $M = \operatorname{diag}\big(m\|a_i\|_2^2\big)^{-1}$.



Simultaneous Iterative Reconstruction Techniques (SIRT)

A general class of methods:

$$x^{k+1} = x^k + \omega\, D\, A^T M\, (b - A\, x^k), \qquad k = 0, 1, 2, \ldots$$

	D	M
Landweber	I	I
Projected gradient descent		
Cimmino	I	$\frac{1}{m}\,\mathrm{diag}\left(\frac{1}{\|a_i\|_2^2}\right)$
Landweber with row normalization		
CAV	I	$\mathrm{diag}\left(\frac{1}{\|a_i\|_S^2}\right)$
Component Averaging		$S = \mathrm{diag}(\mathtt{nnz}(\text{column } j))$
DROP	S^{-1}	$\mathrm{diag}\left(\frac{1}{\|a_i\|_2^2}\right)$
Diagonally relaxed orthogonal projection		
SART	$\mathrm{diag}\left(\frac{1}{\|a^j\|_1}\right)$	$\mathrm{diag}\left(\frac{1}{\|a_i\|_1}\right)$
Simultaneous algebraic reconstruction technique		

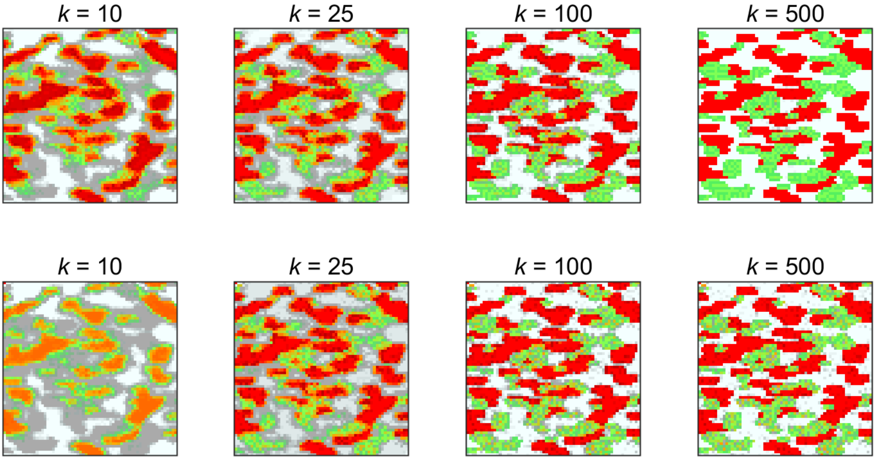
Notation: $\quad a_i = A(i,:) = \text{row}, \qquad a^j = A(:,j) = \text{column}.$

Example of Cimmino – Test Problems

Image size 128×128.   Data: 360 projection angles in $[0^\circ, 360^\circ]$,
181 detector pixels (181 rays per projection), 2 % Gaussian noise.

We use a synthetic problem $A\,x = b$ with a "phantom" – i.e., a test image – inspired by a colorful Dutch cheese.

Example of Cimmino – Results



Top: no noise. Bottom: 2% Gaussian noise.

Asymptotic Convergence for Kaczmarz's Method

Galántai (2004); Strohmer and Vershynin (2009)

Assume that A is **invertible** and that all rows are scaled to unit 2-norm.

$$\left.\begin{aligned} \|x^\ell - \bar{x}\|_2^2 &\leq \Big(1 - \det(A)^2\Big)^\ell \, \|x^0 - \bar{x}\|_2^2 \\ \mathcal{E}\big(\|x^\ell - \bar{x}\|_2^2\big) &\leq \left(1 - \frac{1}{n\,\kappa^2}\right)^\ell \|x^0 - \bar{x}\|_2^2 \end{aligned}\right\} \quad \ell = 1, 2, \dots\ ,$$

where $\mathcal{E}(\cdot)$ = expected value, $\bar{x} = A^{-1} b$, $\kappa = \|A\|_2 \, \|A^{-1}\|_2$, and ℓ counts the number of row actions. This is linear convergence.

Asymptotic Convergence for Kaczmarz's Method

Galántai (2004); Strohmer and Vershynin (2009)

Assume that A is **invertible** and that all rows are scaled to unit 2-norm.

$$\left.\begin{aligned} \|x^\ell - \bar{x}\|_2^2 &\leq \left(1 - \det(A)^2\right)^\ell \|x^0 - \bar{x}\|_2^2 \\ \mathcal{E}\left(\|x^\ell - \bar{x}\|_2^2\right) &\leq \left(1 - \frac{1}{n\,\kappa^2}\right)^\ell \|x^0 - \bar{x}\|_2^2 \end{aligned}\right\} \quad \ell = 1, 2, \ldots \ ,$$

where $\mathcal{E}(\cdot) =$ expected value, $\bar{x} = A^{-1}b$, $\kappa = \|A\|_2 \, \|A^{-1}\|_2$, and ℓ counts the number of row actions. This is linear convergence.

When κ is large we have

$$\left(1 - \frac{1}{n\,\kappa^2}\right)^\ell \approx 1 - \frac{\ell}{n\,\kappa^2}.$$

After $\ell = n$ updates, i.e., one full sweep, the reduction factor is $1 - 1/\kappa^2$.

Asymptotic Convergence for Cimmino (a SIRT Method)

Follows from Nesterov (2004)

Assume that A is invertible and scaled such that $\|A\|_2^2 = m$.

$$\|x^k - \bar{x}\|_2^2 \leq \left(1 - \frac{2}{1+\kappa^2}\right)^k \|x^0 - \bar{x}\|_2^2$$

where $\bar{x} = A^{-1}b$ and $\kappa = \|A\|_2 \, \|A^{-1}\|_2$. Again: linear convergence.

Asymptotic Convergence for Cimmino (a SIRT Method)

Follows from Nesterov (2004)

Assume that A is invertible and scaled such that $\|A\|_2^2 = m$.

$$\|x^k - \bar{x}\|_2^2 \leq \left(1 - \frac{2}{1+\kappa^2}\right)^k \|x^0 - \bar{x}\|_2^2$$

where $\bar{x} = A^{-1}b$ and $\kappa = \|A\|_2 \, \|A^{-1}\|_2$. Again: linear convergence.

When κ is large then we have the approximate upper bound

$$\|x^k - \bar{x}\|_2^2 \lesssim (1 - 2/\kappa^2)^k \, \|x^0 - \bar{x}\|_2^2,$$

showing that in each iteration the error is reduced by a factor $1 - 2/\kappa^2$.

Almost the same factor as in one full sweep in Kaczmarz's method.

Real Problems Have Noisy Data

A standard topic in numerical linear algebra: solve $A\,x = b$.

Don't do this for inverse problems with noisy data!

Real Problems Have Noisy Data

A standard topic in numerical linear algebra: solve $A\,x = b$.

Don't do this for inverse problems with noisy data!

The right-hand side b (the data) is a sum of noise-free data $\bar{b} = A\,\bar{x}$ from the ground-truth image $\bar{x}$ plus a noise component e:

$$b = A\,\bar{x} + e, \qquad \bar{x} = \text{ground truth}, \qquad e = \text{noise}.$$

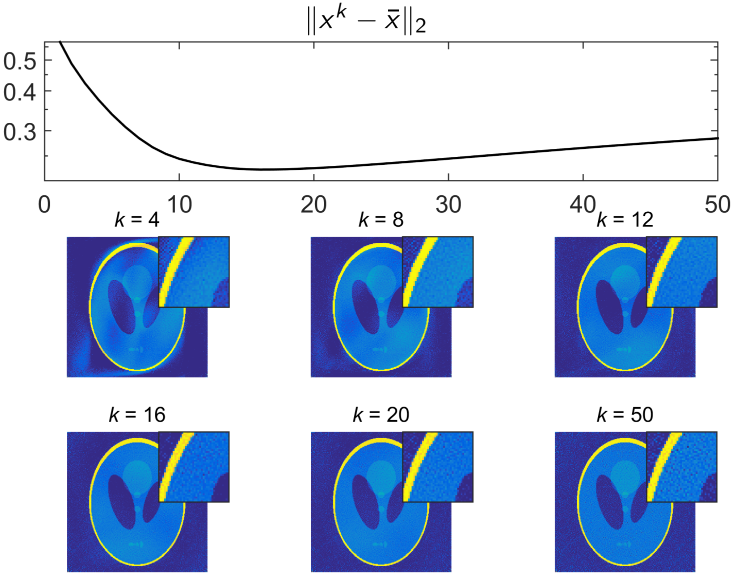
The naïve solution $x^{\text{naïve}} = A^{-1}b$ is undesired, because it has a large component coming from the noise in the data:

$$x^{\text{naïve}} = A^{-1}b = A^{-1}(A\,\bar{x} + e) = \bar{x} + A^{-1}e.$$

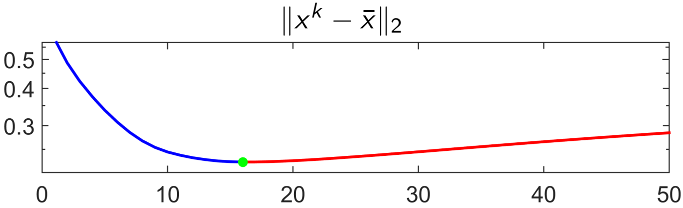
The component $A^{-1}e$ dominates over $\bar{x}$, because A is ill conditioned.

But something interesting happens during the iterations ...

The Reconstruction Error for Kaczmarz's Method



Semi-Convergence



- In the initial iterations x^k approaches the unknown ground truth $\bar{x}$.
- During later iterations x^k converges to the undesired $x^{\text{naïve}} = A^{-1}b$.
- Stop the iterations when the convergence behavior changes.

Then we achieve a regularized solution: an approximation to the noise-free solution which is not too perturbed by the noise in the data.

- Today we explain *why* we have semi-convergence for noisy data.
- How to stop the iterations at the right time is a *different story*.

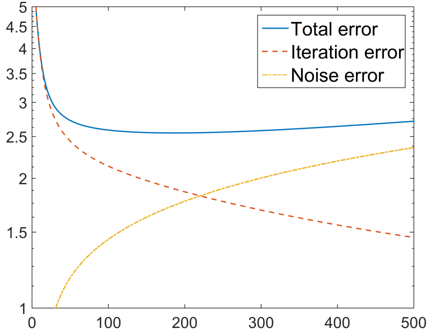
Convergence Analysis: Split the Error

Let $\bar{x}^k$ denote the iterates for a noise-free right-hand side. We consider:

$$\underbrace{x^k - \bar{x}}_{\text{total error}} = \underbrace{x^k - \bar{x}^k}_{\text{noise error}} + \underbrace{\bar{x}^k - \bar{x}}_{\text{iteration error}}$$

We expect the iteration error to decrease and the noise error to increase.

Then we have *semi-convergence*, when the noise error starts to dominate:



Analysis of Semi-Convergence for Cimmino

Consider Cimmino's method, and use the SVD $\boxed{M^{\frac{1}{2}} A = \sum_{i=1}^{n} u_i \, \sigma_i \, v_i^T}$

Van der Sluis & Van der Vorst (1990)

The iterate x^k is a **filtered SVD solution:**

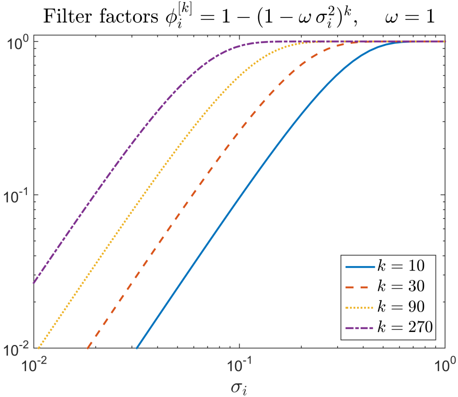
$$x^k = \sum_{i=1}^{n} \varphi_i^{[k]} \, \frac{u_i^T (M^{\frac{1}{2}} b)}{\sigma_i} \, v_i , \qquad \varphi_i^{[k]} = 1 - \left(1 - \omega \, \sigma_i^2\right)^k .$$

Recall that we solve *noisy* systems $A\,x = b$ with $b = A\,\bar{x} + e$. Then:

$$x^k - \bar{x} = \underbrace{\sum_{i=1}^{n} \varphi_i^{[k]} \, \frac{u_i^T (M^{\frac{1}{2}} e)}{\sigma_i} \, v_i}_{\text{noise error}} - \underbrace{\sum_{i=1}^{n} (1 - \varphi_i^{[k]}) \, v_i^T \, \bar{x} \, v_i}_{\text{iteration error}} .$$

Fact: the iteration error decreases. Aim: show that noise error increases.

The Behavior of the Filter Factors



The iteration error $\sum_{i=1}^{n} (1 - \varphi_i^{[k]})\, v_i^T\, \bar{x}\, v_i$ decreases monotonically.

The filter factors *dampen* the "inverted noise" in $\sum_{i=1}^{n} \varphi_i^{[k]} \frac{u_i^T (M^{\frac{1}{2}} e)}{\sigma_i}$.

Note: $\omega\,\sigma_i^2 \ll 1 \Rightarrow \varphi_i^{[k]} \approx k\,\omega\,\sigma_i^2$ showing that k and ω play the same role.

The Spectral Behavior of the Noise Error

Recall: the noise error =

$\sum_{i=1}^{n} \frac{\varphi_i^{[k]}}{\sigma_i} u_i^T (M^{\frac{1}{2}} e) v_i$

and v_i is a *spectral basis:*

▷ large $\sigma_i \sim$ low-freq. v_i

▷ small $\sigma_i \sim$ high-freq. v_i

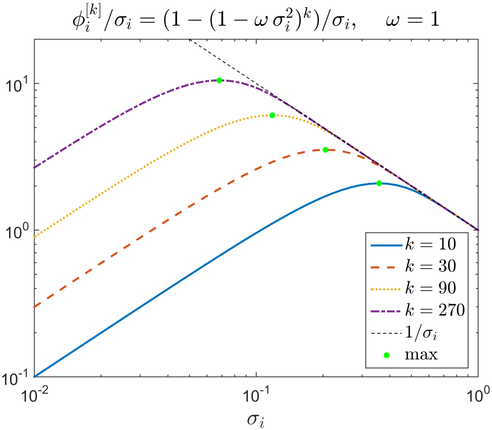
The Spectral Behavior of the Noise Error

Recall: the noise error = $\sum_{i=1}^{n} \frac{\varphi_i^{[k]}}{\sigma_i}\, u_i^T (M^{\frac{1}{2}} e)\, v_i$

and v_i is a *spectral basis:*

▷ large $\sigma_i \sim$ low-freq. v_i

▷ small $\sigma_i \sim$ high-freq. v_i



- Each curve has a maximum for $\sigma_i \approx 1.12/\sqrt{k\,\omega}$.
- As k increases, more noise is included and the SVD-spectrum changes.
- As k increases, the noise error gets dominated by higher frequencies.

Constrained Problems

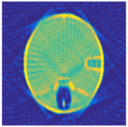
In many applications we can improved the reconstruction by including simple constraints:

$$\boxed{\min_x \|A\,x - b\|_2 \quad \text{s.t.} \quad x \in \mathcal{C}}$$

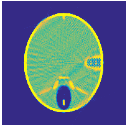
where $\mathcal{C}$ is a convex set, e.g.,

- $\mathcal{C} = \mathbb{R}^n$ – nonnegativity constraints.
- $\mathcal{C} = [0,1]^n$ – box constraints.

No constr.



Box constr.



Constrained Problems

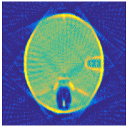
In many applications we can improved the reconstruction by including simple constraints:

$$\boxed{\min_x \|A\,x - b\|_2 \quad \text{s.t.} \quad x \in \mathcal{C}}$$

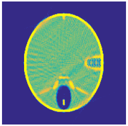
where $\mathcal{C}$ is a convex set, e.g.,

- $\mathcal{C} = \mathbb{R}^n$ – nonnegativity constraints.
- $\mathcal{C} = [0,1]^n$ – box constraints.

No constr.



Box constr.



Let $\mathcal{P}_\mathcal{C}$ denote the orthogonal projector on $\mathcal{C}$.

Kaczmarz (ART) with projection:

$$x \leftarrow \mathcal{P}_\mathcal{C}\left(x + \omega\,\frac{b_i - a_i^T x}{\|a_i\|_2^2}\,a_i\right) \;, \qquad i = 1, 2, 3, \ldots$$

SIRT with projection:

$$x^{k+1} = \mathcal{P}_\mathcal{C}\Big(x^k + \omega\,D\,A^T M\,(b - A\,x^k)\Big)\,, \qquad k = 0, 1, 2, \ldots$$

Analysis of Semi-Convergence for Projected Cimmino

For *constrained* problems we cannot perform an SVD analysis.

Let $\bar{x}$ be the constrained solution to the noise-free problem:

$$\bar{x} = \mathrm{argmin}_{x \in \mathcal{C}} \|A\,x - \bar{b}\|_M, \qquad \bar{b} = A\,\bar{x} = \text{pure data}$$

and let $\bar{x}^k$ denote the iterates when applying Projected Cimmino to $\bar{b}$.

Then we consider an norm-wise analysis

$$\|x^k - \bar{x}\|_2 \leq \underbrace{\|x^k - \bar{x}^k\|_2}_{\text{noise error}} + \underbrace{\|\bar{x}^k - \bar{x}\|_2}_{\text{iteration error}} .$$

Analysis of Semi-Convergence for Projected Cimmino

For *constrained* problems we cannot perform an SVD analysis.

Let $\bar{x}$ be the constrained solution to the noise-free problem:

$$\bar{x} = \operatorname{argmin}_{x \in \mathcal{C}} \|A\,x - \bar{b}\|_M, \qquad \bar{b} = A\,\bar{x} = \text{pure data}$$

and let $\bar{x}^k$ denote the iterates when applying Projected Cimmino to $\bar{b}$.

Then we consider an norm-wise analysis

$$\|x^k - \bar{x}\|_2 \leq \underbrace{\|x^k - \bar{x}^k\|_2}_{\text{noise error}} + \underbrace{\|\bar{x}^k - \bar{x}\|_2}_{\text{iteration error}} \quad .$$

We already considered the decreasing iteration error:

$$\|\bar{x}^k - \bar{x}\|_2 \lesssim (1 - 2/\kappa^2)^k \, \|x^0 - \bar{x}\|_2^2 \ .$$

Now we must consider the noise error (which we expect to grow with k).

The Noise Error for Projected Cimmino

Elfving, H, Nikazad (2012)

The noise error in Projected Cimmino is bounded by

$$\|x^k - \bar{x}^k\|_2 \leq \frac{\sigma_1}{\sigma_n} \frac{1-(1-\omega\,\sigma_n^2)^k}{\sigma_n} \|M^{\frac{1}{2}} e\|_2 \ .$$

As long as $\omega\,\sigma_n^2 \ll 1$ we have $1-(1-\omega\,\sigma_n^2)^k \approx k\,\omega\,\sigma_n^2$ and thus

$$\boxed{\|x^k - \bar{x}^k\|_2 \lesssim \omega\, k\, \sigma_1 \|M^{\frac{1}{2}} e\|_2 \ ,}$$

showing again that k and ω play the same role in the error bound.

Analysis of Semi-Convergence for ART – Setting the Stage

Elfving, Nikazad (2009)

A **full sweep** of ART can be written in a form that *resembles* SIRT:

$$x^{k+1} = x^k + \omega\, A^T \widehat{M}\, (b - A\, x^k)\ , \qquad \widehat{M} = (\Delta + \omega\, L)^{-1}\ .$$

where the **nonsymmetric** $\widehat{M}$ comes from the splitting:

$$A\, A^T = L + \Delta + L^T, \qquad \Delta = \mathrm{diag}(\|a_i\|_2^2)\ ,$$

and where L is strictly lower triangular.

Analysis of Semi-Convergence for ART – Setting the Stage

Elfving, Nikazad (2009)

A **full sweep** of ART can be written in a form that *resembles* SIRT:

$$x^{k+1} = x^k + \omega\, A^T \widehat{M}\,(b - A\,x^k)\ , \qquad \widehat{M} = (\Delta + \omega\, L)^{-1}\ .$$

where the **nonsymmetric** $\widehat{M}$ comes from the splitting:

$$A\,A^T = L + \Delta + L^T, \qquad \Delta = \mathrm{diag}(\|a_i\|_2^2)\ ,$$

and where L is strictly lower triangular.

Simple manipulations show that the noise error is given by

$$\begin{aligned} x^k - \bar{x}^k &= (I - \omega\, A^T \widehat{M}\, A)\,(x^{k-1} - \bar{x}^{k-1}) + \omega\, A^T \widehat{M}\, e \\ &= \omega \sum_{j=1}^{k-1} (I - \omega\, A^T \widehat{M}\, A)^j\, A^T \widehat{M}\, e\ . \end{aligned}$$

Analysis of Semi-Convergence for ART – Results

Elfving, H, Nikazad (2014)

Let $\delta = \|A^T \widehat{M}\, e\|_2$ and σ_r = smallest nonzero singular value of A.

We obtain a bound which resembles that of Cimmino:

$$\boxed{\|x^k - \bar{x}^k\|_2 \le \omega\, k\, \delta + O(\sigma_r^2)}$$

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

$$\boxed{\|x^k - \bar{x}^k\|_2 \le \frac{\sqrt{\omega}}{\sigma_r} \sqrt{k}\, \delta + O(\sigma_r^2)}$$

These results also hold for constrained problems, provided that

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

Numerical Results – Parallel-Beam X-Ray Tomography

Test problem

- ▷ 200×200 phantom
- ▷ 60 projections at
- ▷ $3^\circ, 6^\circ, 9^\circ, \ldots, 180^\circ$
- ▷ $m = 15\,232$
- ▷ $n = 40\,000$

Numerical Results – Parallel-Beam X-Ray Tomography

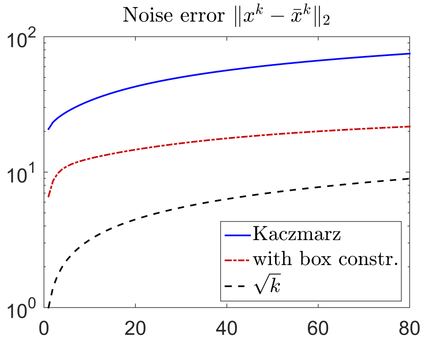
Test problem

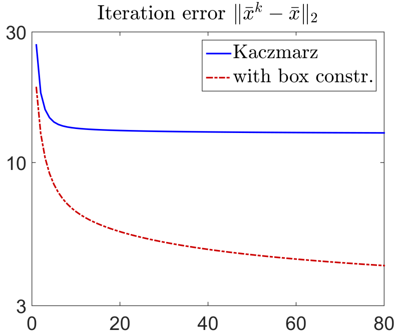
- 200 × 200 phantom
- 60 projections at
- $3^\circ, 6^\circ, 9^\circ, \ldots, 180^\circ$
- $m = 15\,232$
- $n = 40\,000$

The upper bound. We estimate

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

Our bound $\frac{\sqrt{\omega}}{\sigma_r}\delta\sqrt{k}$ is a huge over-estimate; the factor $\sqrt{k}$ correctly *tracks* the noise error.





The Story So Far

Interesting stuff not covered here:

- Convergence of column-action methods.
- Connections to first-order optimization methods.
- Pre-asymptotic convergence of ART; Jiao, Jin, Lu (2017).
- Choice of relaxation parameters; stopping rules.

The Story So Far

Interesting stuff not covered here:

- Convergence of column-action methods.
- Connections to first-order optimization methods.
- Pre-asymptotic convergence of ART; Jiao, Jin, Lu (2017).
- Choice of relaxation parameters; stopping rules.

What we did cover:

- Review of the convergence for noise-free data (iteration error).

The Story So Far

Interesting stuff not covered here:

- Convergence of column-action methods.
- Connections to first-order optimization methods.
- Pre-asymptotic convergence of ART; Jiao, Jin, Lu (2017).
- Choice of relaxation parameters; stopping rules.

What we did cover:

- Review of the convergence for noise-free data (iteration error).
- Illustration of semi-convergence.

The Story So Far

Interesting stuff not covered here:

- Convergence of column-action methods.
- Connections to first-order optimization methods.
- Pre-asymptotic convergence of ART; Jiao, Jin, Lu (2017).
- Choice of relaxation parameters; stopping rules.

What we did cover:

- Review of the convergence for noise-free data (iteration error).
- Illustration of semi-convergence.
- Recent convergence results (upper bounds) for the noise error.

The Story So Far

Interesting stuff not covered here:

- Convergence of column-action methods.
- Connections to first-order optimization methods.
- Pre-asymptotic convergence of ART; Jiao, Jin, Lu (2017).
- Choice of relaxation parameters; stopping rules.

What we did cover:

- Review of the convergence for noise-free data (iteration error).
- Illustration of semi-convergence.
- Recent convergence results (upper bounds) for the noise error.

Unmatched Projector/Backprojector Pairs

Recall the basic iteration: $x^{k+1} = x^k + \omega A^T M (b - A x^k)$.

We take notation literally – the backprojector A^T is really the transposed of the projector A. *Otherwise the theory and the algorithms do not work.*

Unmatched Projector/Backprojector Pairs

Recall the basic iteration: $x^{k+1} = x^k + \omega A^T M\,(b - A\,x^k)$.

We take notation literally – the backprojector A^T is really the transposed of the projector A. *Otherwise the theory and the algorithms do not work.*

But many software packages implement the backprojector in such a way that it is **not** the exact transposed of the projector.

Unmatched Projector/Backprojector Pairs

Recall the basic iteration: $x^{k+1} = x^k + \omega A^T M\,(b - A\,x^k)$.

We take notation literally – the backprojector A^T is really the transposed of the projector A. *Otherwise the theory and the algorithms do not work.*

But many software packages implement the backprojector in such a way that it is **not** the exact transposed of the projector.

1. Philosophy: different discretization schemes may be appropriate for projection and backprojection.

Unmatched Projector/Backprojector Pairs

Recall the basic iteration: $x^{k+1} = x^k + \omega A^T M (b - A x^k)$.

We take notation literally – the backprojector A^T is really the transposed of the projector A. *Otherwise the theory and the algorithms do not work.*

But many software packages implement the backprojector in such a way that it is **not** the exact transposed of the projector.

1. Philosophy: different discretization schemes may be appropriate for projection and backprojection.
2. Practicality: HPC software should make the *most efficient use* of multi-core processors, GPUs and other hardware accelerators.

Unmatched Projector/Backprojector Pairs

Recall the basic iteration: $x^{k+1} = x^k + \omega A^T M\,(b - A\,x^k)$.

We take notation literally – the backprojector A^T is really the transposed of the projector A. *Otherwise the theory and the algorithms do not work.*

But many software packages implement the backprojector in such a way that it is **not** the exact transposed of the projector.

1. Philosophy: different discretization schemes may be appropriate for projection and backprojection.
2. Practicality: HPC software should make the *most efficient use* of multi-core processors, GPUs and other hardware accelerators.

We must study the influence of <u>unmatched</u> projector/backprojector pairs on the computed solutions and the convergence of the iterations.

Perturbation Theory for Unmatched Normal Equations

Let $\{A, A^T, \bar{b}\}$ be the unperturbed data, and consider the perturbations

$$\tilde{A} = A + E_A, \qquad \hat{A}^T = A^T + E_{A^T}, \qquad b = \bar{b} + e.$$

Also let $\bar{x}$ denote the unperturbed solution to $A^T A \, \bar{x} = A^T \bar{b}$.

Perturbation Theory for Unmatched Normal Equations

Let $\{A, A^T, \bar{b}\}$ be the unperturbed data, and consider the perturbations

$$\tilde{A} = A + E_A, \qquad \hat{A}^T = A^T + E_{A^T}, \qquad b = \bar{b} + e.$$

Also let $\bar{x}$ denote the unperturbed solution to $A^T A\,\bar{x} = A^T \bar{b}$.

Elfving, H (2018)

When we use the perturbed triple $\{\tilde{A}, \hat{A}^T, b\}$ then we aim at solving the *unmatched normal equations:*

$$\hat{A}^T \tilde{A}\,(\bar{x} + \delta x) = \hat{A}^T b.$$

Omitting higher-order terms, we obtain:

$$\boxed{\|\delta x\|_2 \lesssim \frac{1}{\sigma_r}\big(\|\mathcal{P}_{\mathcal{R}(A)}\, e\|_2 + \|E_A \bar{x}\|_2\big) + \frac{1}{\sigma_r^2}\|E_{A^T}(\bar{b} - A\,\bar{x})\|_2}$$

For inconsistent systems, the solution is more sensitive to E_{A^T} than E_A.

Convergence Analysis for Unmatched Pairs

To set the stage we consider the generic **BA Iteration**

$$\boxed{x^{k+1} = x^k + \omega\, B\,(b - A\,x^k)\ , \qquad \omega > 0}$$

Generally not related to solving a minimization problem!

Convergence Analysis for Unmatched Pairs

To set the stage we consider the generic **BA Iteration**

$$\boxed{x^{k+1} = x^k + \omega\, B\,(b - A\,x^k)\ , \qquad \omega > 0}$$

Generally not related to solving a minimization problem!

It is a *fixed-point iteration* whose convergence depends on the product BA.

- Any fixed point x^* satisfies the unmatched normal eq. $B\,A\,x^* = B\,b$.
- If BA is invertible then $x^* = (BA)^{-1}B\,b$.

Convergence Analysis for Unmatched Pairs

To set the stage we consider the generic **BA Iteration**

$$\boxed{x^{k+1} = x^k + \omega\, B\,(b - A\,x^k)\ , \qquad \omega > 0}$$

Generally not related to solving a minimization problem!

It is a *fixed-point iteration* whose convergence depends on the product BA.

- Any fixed point x^* satisfies the unmatched normal eq. $BA\,x^* = B\,b$.
- If BA is invertible then $x^* = (BA)^{-1}B\,b$.

Shi, Wei, Zhang (2011); Elfving, H (2018)

The **BA Iteration** converges to a solution of $BA\,x = B\,b$ if and only if

$$0 < \omega < \frac{2\,\mathrm{Re}(\lambda_j)}{|\lambda_j|^2} \quad \text{and} \quad \mathrm{Re}(\lambda_j) > 0, \qquad \{\lambda_j\} = \mathtt{eig}(BA)\ .$$

Zeng & Gullberg (2000): similar analysis but ignoring complex λ_j.

More Convergence Results for Unmatched Pairs

Dong, H, Hochstenbach, Riis (2019) – for the nerds

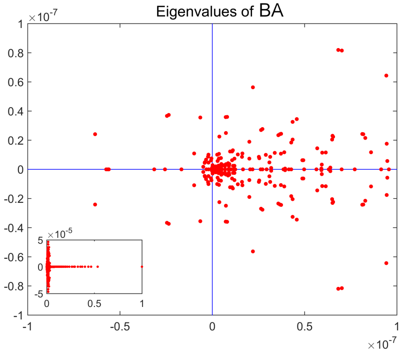
The following requirements for a *unique fixed point* are equivalent:

1. $BA : \mathcal{R}(B) \to \mathcal{R}(B)$ is nonsingular.
2. For every $b \in \mathbb{R}^m$, $BA\,x = B\,b$ has a unique solution $x \in \mathcal{R}(B)$.
3. $\mathcal{R}(B) \cap \mathcal{N}(BA) = \{0\}$.
4. $\mathcal{N}(BA\,B) = \mathcal{N}(B)$.
5. $\mathcal{R}(BA\,B) = \mathcal{R}(B)$.
6. $\mathrm{rank}(BA\,B) = \mathrm{rank}(B)$.
7. A is nonsingular on $\mathcal{R}(B)$ and B is nonsingular on $\mathcal{R}(AB)$.
8. $\mathcal{R}(B) \cap \mathcal{N}(A) = \{0\}$ and $\mathcal{R}(AB) \cap \mathcal{N}(B) = \{0\}$.

Here $\mathcal{R}(\cdot)$ = range and $\mathcal{N}(\cdot)$ = null space.

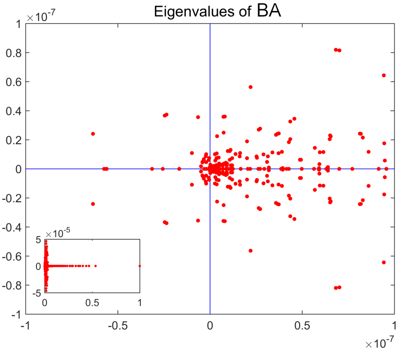
Numerical Example (no Noise) with Negative Real Parts

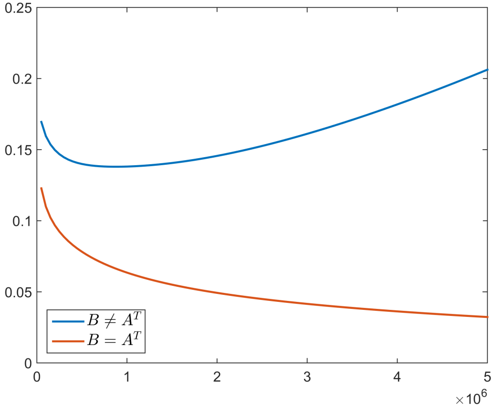
Parallel-beam CT, unmatched pair from *ASTRA*, 64×64 Shepp-Logan phantom, 90 proj. angles, 60 detector pixels, $\min \mathrm{Re}(\lambda_j) = -6.4 \cdot 10^{-8}$.



Numerical Example (no Noise) with Negative Real Parts

Parallel-beam CT, unmatched pair from *ASTRA*, 64×64 Shepp-Logan phantom, 90 proj. angles, 60 detector pixels, $\min \mathrm{Re}(\lambda_j) = -6.4 \cdot 10^{-8}$.





For now we assume that $\mathrm{Re}(\lambda_j) > 0 \ \forall\, j$.

Iteration Error for BA Iteration

For simplicity assume that $\mathcal{N}(BA) = \emptyset \;\Rightarrow\;$ the convergence criterion becomes $\rho(T) < 1$ with $T = I - \omega\, BA$ (otherwise: see paper).

Iteration Error for BA Iteration

For simplicity assume that $\mathcal{N}(BA) = \emptyset \;\Rightarrow\;$ the convergence criterion becomes $\rho(T) < 1$ with $T = I - \omega\, BA$ (otherwise: see paper).

Elfving, H (2018)

The *iteration error* is given by

$$\bar{x}^k - \bar{x} = T^k(\bar{x}^0 - \bar{x}), \qquad \bar{x}^0 = \text{initial vector},$$

and it follows that

$$\|\bar{x}^k - \bar{x}\|_2 \;\le\; \|T^k\|_2\, \|\bar{x}^0 - \bar{x}\|_2 \;\le\; \|T\|_2^k\, \|\bar{x}^0 - \bar{x}\|_2 .$$

In general we cannot assume $\|T\|_2 < 1$; but asymptotically the convergence rate depends on the spectral radius because

$$\lim_{j\to\infty} \|T^j\| = \lim_{j\to\infty} \rho(T^j) = 0,$$

so the convergence rate is *linear*.

Noise Error for BA Iteration

Recall that the *noise error* $x^k - \bar{x}^k$ reveals how the errors e in the right-hand side propagate during the iterations.

From the definition of the **BA Iteration** it follows that

$$x^k - \bar{x}^k = (I - \omega\, BA)\,(x^{k-1} - \bar{x}^{k-1}) + \omega B\, e,$$

and hence by induction, and assuming $x^0 = \bar{x}^0$, it follows that

$$x^k - \bar{x}^k = S_k e \quad \text{with} \quad S_k = \omega \sum_{j=0}^{k-1} (I - \omega\, BA)^j B.$$

Noise Error for BA Iteration

Recall that the *noise error* $x^k - \bar{x}^k$ reveals how the errors e in the right-hand side propagate during the iterations.

From the definition of the **BA Iteration** it follows that

$$x^k - \bar{x}^k = (I - \omega\, BA)\,(x^{k-1} - \bar{x}^{k-1}) + \omega B\, e,$$

and hence by induction, and assuming $x^0 = \bar{x}^0$, it follows that

$$x^k - \bar{x}^k = S_k e \quad \text{with} \quad S_k = \omega \sum_{j=0}^{k-1} (I - \omega\, BA)^j B.$$

Elfving, H (2018)

Similar to iterations with a matched transpose, with $b = A\,\bar{x} + e$ we have

$$\boxed{\|x^k - \bar{x}^k\|_2 \;\leq\; (\omega\, c_{\mathrm{BA}} \|B\|_2)\, k\, \|e\|_2}$$

where we define the constant c_{BA} by: $\sup_j \|(I - \omega\, BA)^j\|_2 \leq c_{\mathrm{BA}}$.

Numerical Experiments – the Influence of Unmatching

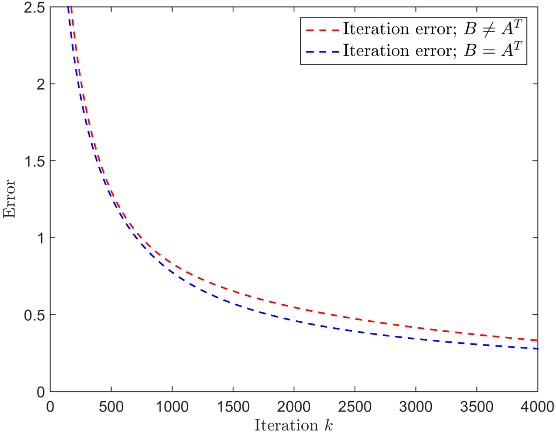
- 64×64 image, 180 proj., 91 detector pixels, A is $16,380 \times 4,096$.
- Unmatched transpose $\hat{A}^T$: generated from A^T by neglecting the smallest 50% of the nonzeros; then $\|E_{A^T}\|_\mathsf{F}/\|A\|_\mathsf{F} = 0.406$.
- Noisy $b = \bar{b} + e$: Gaussian white noise with $\|e\|_2/\|\bar{b}\|_2 = 0.01$.
- Both A and $\hat{A}$ have full rank.
- All real parts of the eigenvalues of $C = \hat{A}^T A$ are positive (the smallest real part is $9.35 \cdot 10^{-7}$).
- For the unperturbed right-hand side $\bar{b} = A\,\bar{x}$, the **BA Iteration** with both $B = A^T$ and $B = \hat{A}^T$ converges to $\bar{x}$.
- For the perturbed right-hand side b, the iteration converges to $\bar{x}$ when $B = A^T$ and to a solution of $\hat{A}^T A\,x = \hat{A}^T b$ when $B = \hat{A}^T$.

We show: $\underbrace{x^k - \bar{x}}_{\text{total error}} = \underbrace{x^k - \bar{x}^k}_{\text{noise error}} + \underbrace{\bar{x}^k - \bar{x}}_{\text{iteration error}}$

Cimmino's method.

Test problem

- ▷ 64×64 phantom
- ▷ 180 projections at
- ▷ $1^\circ, 2^\circ, 3^\circ, \ldots, 180^\circ$
- ▷ $m = 16\,380$
- ▷ $n = 4\,096$

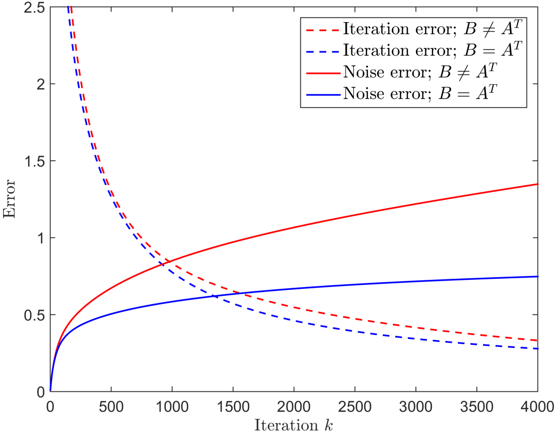


Iteration error: both versions converge to $\bar{x}$; the one with $B \neq A^T$ is slower.

Cimmino's method.

Test problem

- ▷ 64×64 phantom
- ▷ 180 projections at
- ▷ $1°, 2°, 3°, \ldots, 180°$
- ▷ $m = 16\,380$
- ▷ $n = 4\,096$



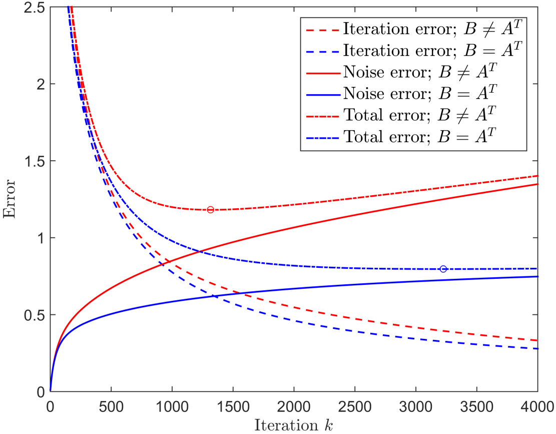
Iteration error: both versions converge to $\bar{x}$; the one with $B \neq A^T$ is slower.

Noise error: the one for $B \neq A^T$ increases faster.

Cimmino's method.

Test problem

- ▷ 64×64 phantom
- ▷ 180 projections at
- ▷ $1^\circ, 2^\circ, 3^\circ, \ldots, 180^\circ$
- ▷ $m = 16\,380$
- ▷ $n = 4\,096$

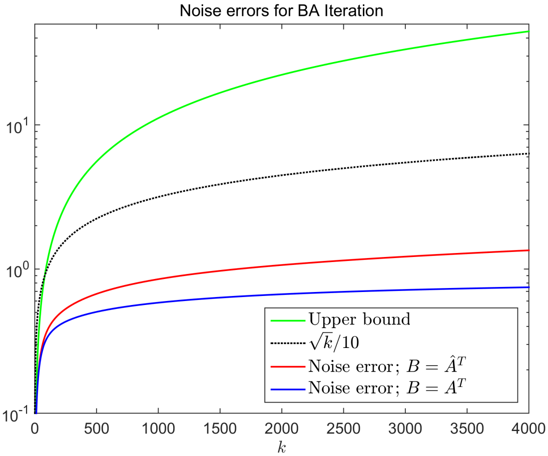


Iteration error: both versions converge to $\bar{x}$; the one with $B \neq A^T$ is slower.

Noise error: the one for $B \neq A^T$ increases faster.

Total error: semi-convergence, the iteration with $B \neq A^T$ reaches the min. error ○ 1.181 after 1314 iterations. This error is 48% larger than the min. error ○ 0.796 for the iterations with A^T, achieved after 3225 iterations.

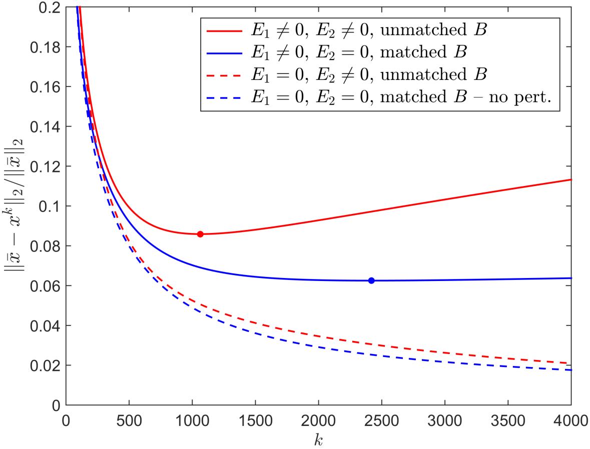
A Closer Look at the Noise Error



General bound: $\|x^k - \bar{x}^k\|_2 \leq (\omega c_T \|B\|_2) k \|e\|_2$; but here the error $\approx \sqrt{k}$.

For row/column action methods with *matched pair* we can show $\sqrt{k}$ bound.

Exact Data ($e = 0$) and Errors in the Matrices



Matrix errors $E_1 \neq 0$ also lead to semi-convergence.

Minimum reconstruction error is larger for an unmatched transpose $E_2 \neq 0$.

Did We Prove Semi-Convergence?

Not really:

- we give an *upper* bound for the noise error;
- this bound increases with k,
- and it seems to track the actual noise error in numerical experiments.

Thus we have justified the observed behavior of

$$\text{total error} = \text{iteration error} + \text{noise error}.$$

Did We Prove Semi-Convergence?

Not really:

- we give an *upper* bound for the noise error;
- this bound increases with k,
- and it seems to track the actual noise error in numerical experiments.

Thus we have justified the observed behavior of

$$\text{total error} = \text{iteration error} + \text{noise error}.$$

But we also need a *lower* bound that increases with k:

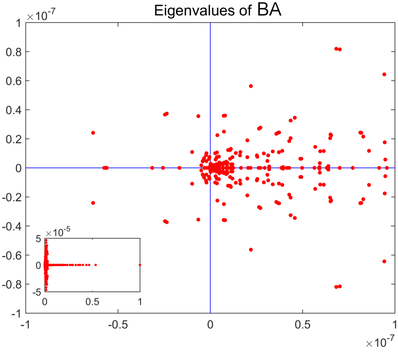
- If the right-hand side error $e \in \mathcal{N}(A)$ then the lower bound is 0 (this is extremely unlikely).
- We (currently) don't know how to derive a nonzero increasing lower bound for the case $e \notin \mathcal{N}(A)$.

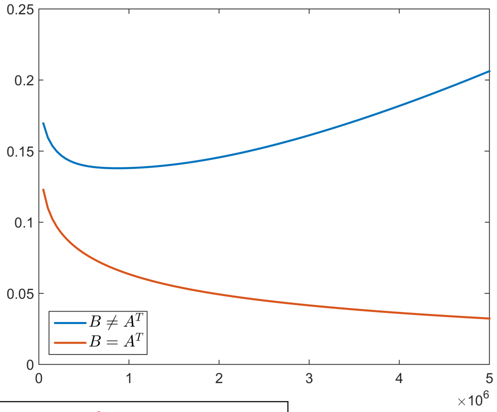
The Story So Far

- We studied the influence of errors in the forward projector and the backprojector.
- Our perturbation analysis shows that the least squares solution is more sensitive to errors in A^T than in A.
- We derived bounds for the errors in the iteration vectors for a generic algorithm that includes many well-known algebraic iterative methods.
- Numerical examples demonstrate that an unmatched matrix pair leads to a less accurate reconstruction than with a matched transpose.
- Next up: "fix" the iterative algorithms when there are eigenvalues with a negative real part.

And Now: Eigenvalues with Negative Real Parts

Parallel-beam CT, unmatched pair from *ASTRA*, 64×64 Shepp-Logan phantom, 90 projection angles, 60 detector pixels, $\min \mathrm{Re}\, \lambda_j = -6.4 \cdot 10^{-8}$.

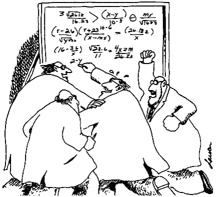




Now we study the case with no asymptotic convergence.

What To Do?

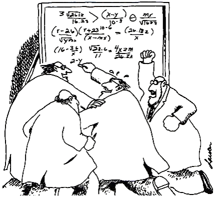
1. Ask the software developers to change their implementation of projection and/or backprojection?
   $\rightarrow$ Significant loss of computational efficiency.
2. Use mathematics to *fix* the nonconvergence.
   $\rightarrow$ What we do here.



What To Do?

1. Ask the software developers to change their implementation of projection and/or backprojection?
   $\rightarrow$ Significant loss of computational efficiency.

2. Use mathematics to *fix* the nonconvergence.
   $\rightarrow$ What we do here.



Take inspiration from the Tikhonov problem

$$\min_x \left\{ \|A\,x - b\|_2^2 + \alpha\,\|x\|_2^2 \right\} \;,$$

for which a gradient step takes the form

$$\begin{aligned} x^{k+1} &= x^k - \omega\,(A^T(b - A\,x) + \alpha\,x^k) \\ &= (1 - \alpha\,\omega)\,x^k + \omega\,A^T(b - A\,x^k)\;. \end{aligned}$$

Note the factor $(1 - \alpha\,\omega)$.

The Shifted BA Iteration


Many thanks to Tommy Elfving for originally suggesting this.

We define the **shifted** version of the BA Iteration:

$$\boxed{x^{k+1} = (1 - \alpha\,\omega)\,x^k + \omega\,B\,(b - A\,x^k)\ , \qquad \omega > 0}$$

with just one extra factor $(1 - \alpha\,\omega)$; simple to implement.

This Shifted BA Iteration is equivalent to applying the BA Iteration with the substitutions

$$A \to \begin{bmatrix} A \\ \sqrt{\alpha}\,I \end{bmatrix}, \qquad B \to \left[\,B\,,\,\sqrt{\alpha}\,I\,\right]\ , \qquad b \to \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

Hence it is "easy" to perform the convergence analysis . . .

Convergence Results

Dong, H, Hochstenbach, Riis (2019)

Let λ_j denote those eigenvalues of BA that are different from $-\alpha$.

Then the Shifted BA Iteration converges to a fixed point if and only if α and ω satisfy

$$0 < \omega < 2\, \frac{\mathrm{Re}\,\lambda_j + \alpha}{|\lambda_j|^2 + \alpha\,(\alpha + 2\,\mathrm{Re}\,\lambda_j)} \qquad \text{and} \qquad \mathrm{Re}\,\lambda_j + \alpha > 0 \ .$$

The fixed point x_α^* satisfies

$$(BA + \alpha\, I)\, x_\alpha^* = B\, b \ .$$

Convergence Results

Dong, H, Hochstenbach, Riis (2019)

Let λ_j denote those eigenvalues of $B\,A$ that are different from $-\alpha$.

Then the Shifted BA Iteration converges to a fixed point if and only if α and ω satisfy

$$0 < \omega < 2\,\frac{\mathrm{Re}\,\lambda_j + \alpha}{|\lambda_j|^2 + \alpha\,(\alpha + 2\,\mathrm{Re}\,\lambda_j)} \qquad \text{and} \qquad \mathrm{Re}\,\lambda_j + \alpha > 0\ .$$

The fixed point x_α^* satisfies

$$(B\,A + \alpha\,I)\,x_\alpha^* = B\,b\ .$$

This result tells us how to choose the shift parameter α:

$$\boxed{\text{Just large enough that } \mathrm{Re}\,\lambda_j + \alpha > 0 \text{ for all } j.}$$

"Perturbation" Result

How much do we perturb the *solution* $\bar{x}^*_\alpha$ – the fixed point – when we introduce $\alpha > 0$?

"Perturbation" Result

How much do we perturb the *solution* $\bar{x}_\alpha^*$ – the fixed point – when we introduce $\alpha > 0$?

Dong, H, Hochstenbach, Riis (2019)

Assume that $BA + \alpha I$ is nonsingular and the right-hand side is noise-free with $b = \bar{b} = A\bar{x}$. Then the corresponding fixed point $\bar{x}_\alpha^*$ satisfies

$$\bar{x} - \bar{x}_\alpha^* = \alpha \, (BA + \alpha \, I)^{-1} \bar{x} \ .$$

Notice the factor α.

"Perturbation" Result

How much do we perturb the *solution* $\bar{x}_\alpha^*$ – the fixed point – when we introduce $\alpha > 0$?

Dong, H, Hochstenbach, Riis (2019)

Assume that $BA + \alpha I$ is nonsingular and the right-hand side is noise-free with $b = \bar{b} = A\bar{x}$. Then the corresponding fixed point $\bar{x}_\alpha^*$ satisfies

$$\bar{x} - \bar{x}_\alpha^* = \alpha (BA + \alpha I)^{-1} \bar{x} \ .$$

Notice the factor α.

With a small α – just large enough to ensure convergence – we compute a slightly perturbed solution (instead of computing nothing).

Eigenvalue Estimates (See Paper for Details)

We need to compute an estimate of the **leftmost** eigenvalue of BA, i.e., the eigenvalue with the minimal real part.

Bring in "Mr. Eigenvalue"
Michiel E. Hochstenbach.

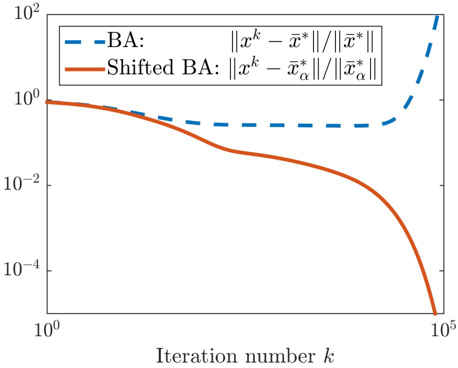
In our paper we discuss five different iterative algorithms:

- Matlab's `eigs(_,_,'smallestreal')` (calls ARPACK): baseline algorithm.
- Algorithms by Meerbergen and coauthors: robust but need too many matrix-vector multiplications.
- Krylov-Schur method by Stewart ($\sim$ implicitly restarted Arnoldi): 30% faster than Matlab's `eigs`.
- Jacobi-Davidson: slower than Krylov-Schur.
- Our own "field-of-values approximation algorithm": competitive with Krylov-Schur.

Numerical Results – Divergence and Convergence

Parallel-beam CT, 128×128 Shepp-Logan phantom, 90 projection angles in $[0^\circ, 180^\circ]$, 80 detector pixels; $m = 7\,200$ and $n = 16\,384$.
Both A and B are generated with the GPU-version of the ASTRA toolbox.

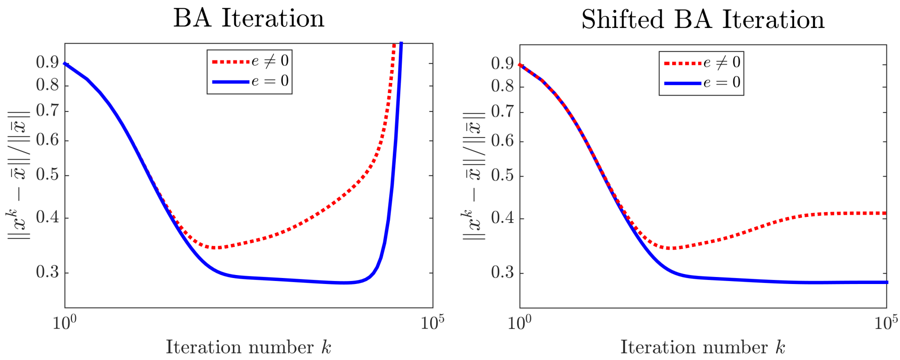
$\rho(BA) = 1.76 \cdot 10^4$
$\alpha = 1.85$



The BA Iteration diverges from $\bar{x}^* = (BA)^{-1} B\,\bar{b}$.

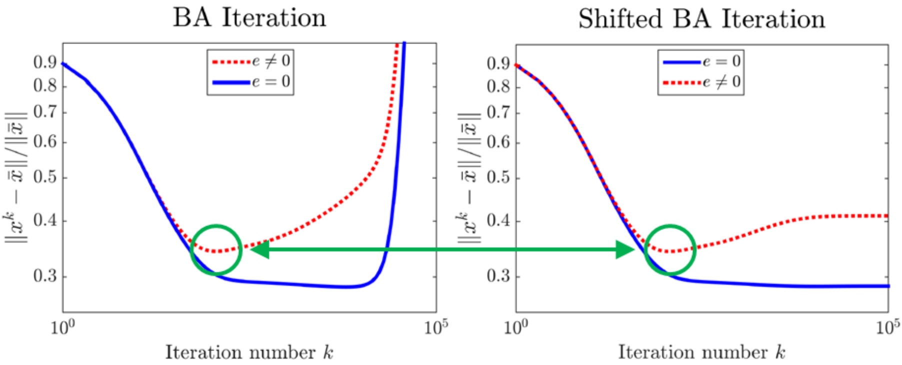
The Shifted BA Iteration converges to fixed point $\bar{x}^*_\alpha = (BA + \alpha\, I)^{-1} B\,\bar{b}$.

Numerical Results – Reconstruction Errors



- The BA Iteration diverges from the ground truth $\bar{x}$.
- The Shifted BA Iteration
  - Without noise: converges to a solution $\bar{x}^*_\alpha$ that approximates $\bar{x}$.
  - With noise: first semi-convergence, then convergence to x^*_α.

Does It Matter?



- For noisy data, the solutions at semi-convergence are almost the same.
- But is this always the case? More research is necessary.
- Also, we prefer iterative methods that converge with or without noise.

Last Part of The Story

- We studied the influence of an <u>unmatched</u> pair of matrices for which backprojection $\neq$ adjoint(projection).
- Focus on SIRT method; also a concern for Kaczmarz-type methods.
- Iterative methods based on unmatched pairs do not solve an optimization problem, but may converge to a fixed point.
- The main criterion for convergence is that all eigenvalues of the iteration matrix must have positive real part.
- If violated, we introduce a small shift that ensures *convergence* to a fixed point that is a *slightly perturbed* solution ($\sim$ Tikhonov).
- The shift is computed via estimation of the leftmost eigenvalue.
- Numerical results confirm our convergence results.

終