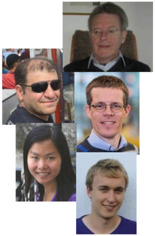
Convergence and Non-Convergence of Algebraic Iterative Reconstruction Methods

Per Christian Hansen
Technical University of Denmark

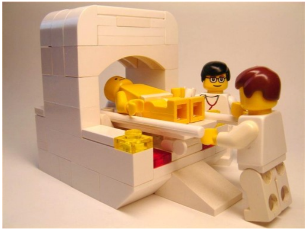


Joint work with

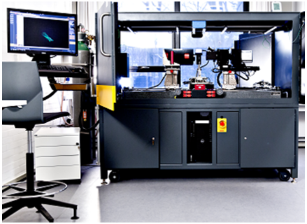
Tommy Elfving – Linköping University
Touraj Nikazad – Iran Univ. Scence & Technology
Michiel E. Hochstenbach – TU Eindhoven
Yiqiu Dong & Nicolai A. B. Riis – DTU Compute

Computed Tomography (CT)

Applications in medicine, materials science, industrial inspection, etc.

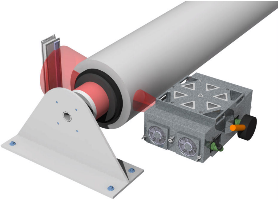


Medical scanner



Lab scanner

Synchrotron



Inspection of pipe

Reconstruction Methods

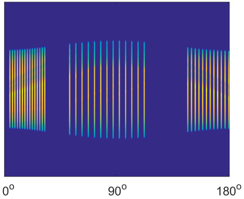
Filtered Back Projection (FBP)

- Very fast, low memory, good results with lots of good data.
- But artifacts appear with noisy and/or limited data.
- Cannot incorporate constraints (e.g., nonnegativity).

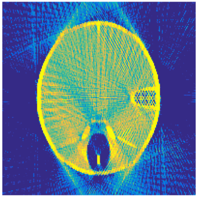
Algebraic Iterative Reconstruction – Kaczmarz, Cimmino, SIRT, etc.

- Very flexible – no assumptions about the CT scanning geometry.
- Easy to incorporate convex constraints (e.g., nonneg./box constraints).
- Often give good reconstructions with noisy and/or limited data.

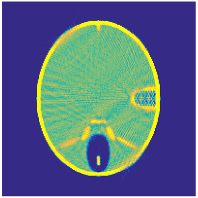
Data (sinogram)



FBP



Kaczmarz



Algebraic Iterative Reconstruction (AIR) Methods

Let $A =$ discretization of forward projection (the Radon transform).

Consider consistent systems:

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

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

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

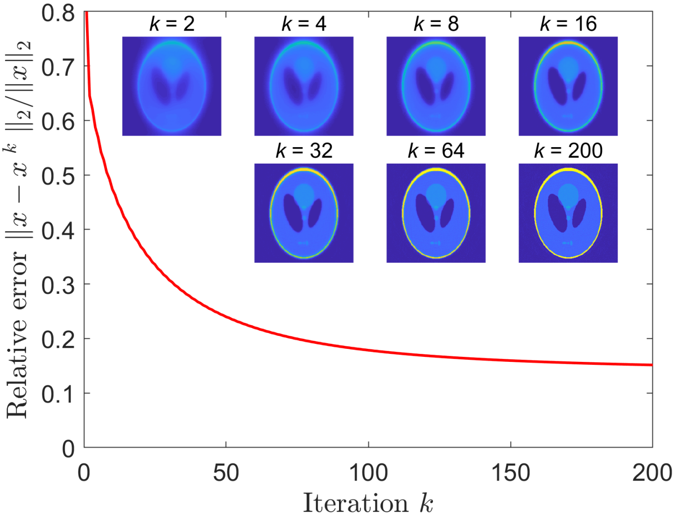
- **Row-action methods** such as Kaczmarz' method

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

- **Column-action methods** = coordinate descent methods

$$x_i^{\ell+1} = x_i^\ell + \omega\,A(:,i)^T\left(b - A(:,i)\,x_i^\ell\right)\ , \qquad i = \ell \bmod n.$$

Example of Iteration Progression



Asymptotic Convergence

Let $\mathcal{E}(\cdot)$ = expected value, $\bar{x} = A^{-1}b$, and $\kappa = \|A\|_2 \, \|A^{-1}\|_2$.

Cimmino's method (simultaneous iterations) – Nesterov (2004)

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

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

Kaczmarz's Method – Galántai (2004); Strohmer and Vershynin (2009)

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

$$\mathcal{E}\left(\|x^\ell - \bar{x}\|_2^2\right) \le \left(1 - \frac{1}{n\,\kappa^2}\right)^\ell \|x^0 - \bar{x}\|_2^2$$

In both cases we have **linear convergence**.

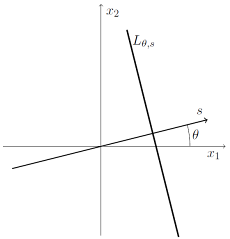
Modelling: Forward and Back Projections

Forward projection $\mathcal{R}$, the Radon transform models the scanner physics via integration of the image f along lines $L_{\theta,s}$

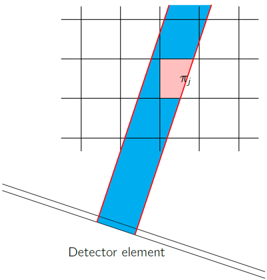
$$\mathcal{R}[f](\theta, s) = \int_{L_{\theta,s}} f(x_1, x_2)\, d\ell = g(\theta, s) = \text{sinogram} \ .$$

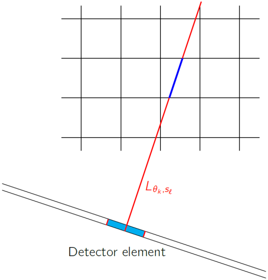
Back projection $\mathcal{B} = \text{adjoint}(\mathcal{R})$, an abstraction, smears g back along $L_{\theta,s}$

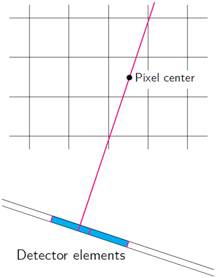
$$\mathcal{B}[g](x_1, x_2) = \int_0^{2\pi} g(\theta, x_1 \cos\theta + x_2 \sin\theta)\, d\theta \ .$$



Discretization: Examples of Models







Forward: strip model	Forward: line model	Back projection model
Reflects the physics	Ray driven	Pixel driven
Not suited for GPUs	Suited for GPUs	Suited for GPUs

Line model: start from detector element centers.
Back projection model: start from image pixel centers.

Unmatched Projectors

Action of forward projector $\mathcal{R}$ $\leftrightsquigarrow$ Multiplication with A.

Action of back projector $\mathcal{B}$ = adjoint($\mathcal{R}$) $\leftrightsquigarrow$ Multiplication with B.

When we can store A then we use A^T for the back projection B.

When A is *too large to store*, we must use matrix-free multiplications of the forward projector and back projector.

HPC software: computational efficiency takes priority $\rightarrow$ $B \neq A^T$.

- Cimmino's method

$$x^{k+1} = x^k + \omega\, B M \,(b - A\, x^k)\ .$$

- Kaczmarz' method

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

What can we say about the convergence of these methods?

Convergence Analysis for Unmatched Pairs

For simplicity we focus on the simple **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 $B\,A$.
Any fixed point x^* satisfies the **unmatched normal equations**

$$B\,A\,x^* = B\,b\ .$$

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

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

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

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

Perturbation Theory

Consider this scenario

$$A = \bar{A} + \delta A \ , \qquad B = \bar{A}^T + \delta B \ , \qquad b = \bar{b} + e$$

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

Using results from Elfving, H (2018)

Write the unmatched normal equations as

$$B\, A\, (\bar{x} + \delta x) = B\, b \ .$$

Omitting higher-order terms, we obtain:

$$\boxed{\|\delta x\|_2 \lesssim \frac{1}{\sigma_r} \big(\|P_{\mathcal{R}(\bar{A})}\, e\|_2 + \|\delta A\, \bar{x}\|_2 \big) + \frac{1}{\sigma_r^2} \|\delta B\, (\bar{b} - \bar{A}\, \bar{x})\|_2}$$

σ_r = smallest nonzero sing. value of $\bar{A}$; $P_{\mathcal{R}(\bar{A})}$ = orthog. proj. on $\mathcal{R}(\bar{A})$.

For inconsistent systems, the solution is more sensitive to δB than δA.

Convergence Analysis: Split the Error

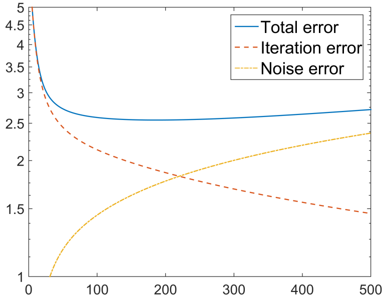
First assume that $\mathrm{Re}(\lambda_j) > 0\ \forall j$ is satisfied, i.e., we have convergence.

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:



Noise Error for BA Iteration

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

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

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

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

Elfving, H (2018)

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

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

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

Numerical Experiments – the Influence of Unmatching

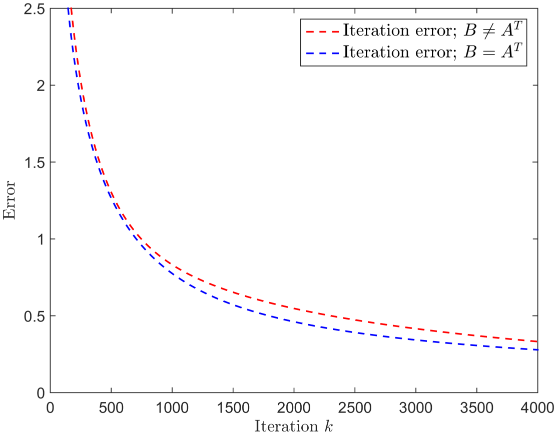
- 64×64 image, 180 proj., 91 detector pixels, A is $16,380 \times 4,096$.
- The unmatched transpose satisfies $\|B - A^T\|_F / \|A\|_F = 0.406$.
- Noisy $b = \bar{b} + e$: Gaussian white noise with $\|e\|_2 / \|\bar{b}\|_2 = 0.01$.
- Both A and B have full rank.
- All real parts of the eigenvalues of BA are positive (the smallest real part is $9.35 \cdot 10^{-7}$).

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

Cimmino's method.

Test problem

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

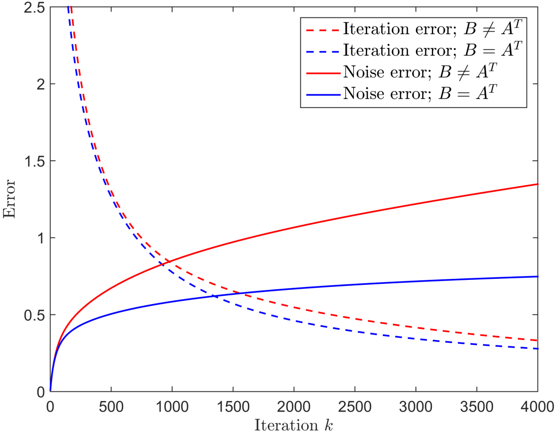


Iteration error: both versions converge to $\bar{x}$; the one with $B \neq A^T$ is slower.

Cimmino's method.

Test problem

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



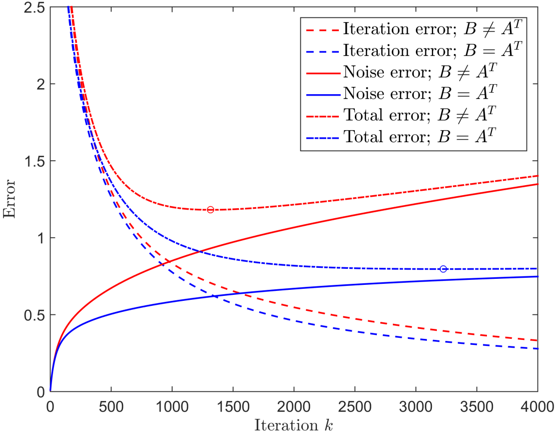
Iteration error: both versions converge to $\bar{x}$; the one with $B \neq A^T$ is slower.

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

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.

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.

How to stop at the point of semi-convergence?

Notation: $\eta^2 = \mathcal{E}(\|e\|_2^2)$ and $t_k = \mathrm{trace}(A\,A_k^\#)$ with $x_k = A_k^\# b$.

1. **Discrepancy principle**

$$\|b - A\,x^k\|_2^2 \approx \eta^2 \ .$$

2. **Unbiased predictive risk estimation**

$$\text{minimize } U_k = \|b - A\,x^k\|_2^2 + 2\,\eta^2\,t_k - \eta^2\,m \ .$$

3. **Generalized cross validation**

$$\text{minimize } G_k = \frac{\|b - A\,x^k\|_2^2}{(m - t_k)} \ .$$

4. **Normalized cumulative periodogram**

   stop when the power spectrum of $b - A\,x^k$ resembles white noise.

But this is another talk.

Did We <u>Prove</u> 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.

This supports 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.

This supports the observed behavior of

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

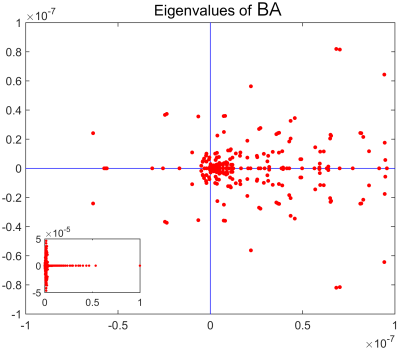
But we also need a *lower* bound for the noise error, that increases with k:

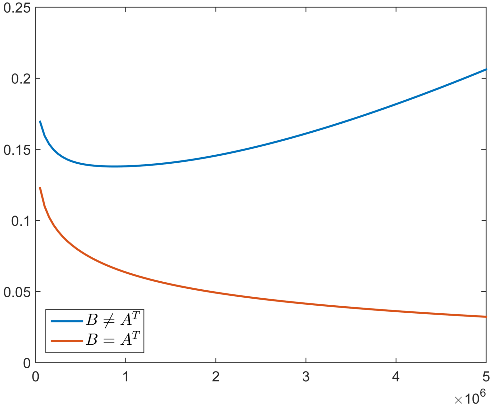
- If the right-hand side error $e \in \mathcal{N}(B)$ then the lower bound is 0 (this is extremely unlikely).
- A lower bound for Kaczmarz's method was derived in van Lith, H, Hochstenbach (2021).
- Currently no known lower bound for the BA Iteration.

Numerical Example (no Noise) with Negative Real Parts

Now we consider the case where some $\mathrm{Re}(\lambda_j) < 0$, i.e., no convergence.

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

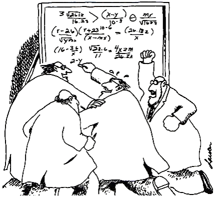




Eigenvalues with Negative Real Parts – What To Do?

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

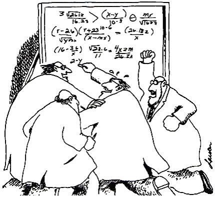
2. Use mathematics to *fix* the nonconvergence.
   $\rightarrow$ What we do here.



Eigenvalues with Negative Real Parts – What To Do?

1. Ask the software developers to change their implementation of forward projection and/or back projection?
   $\rightarrow$ 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

$$A \to \begin{bmatrix} AA \\ \sqrt{\alpha}\,I \end{bmatrix}, \qquad B \to \begin{bmatrix} B\,, \sqrt{\alpha}\,I \end{bmatrix}\ , \qquad b \to \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

Hence it is “easy” to perform the convergence analysis . . .

Convergence Results

Dong, H, Hochstenbach, Riis (2019)

Let λ_j denote the eigenvalues of BA.

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

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

The fixed point x_α^* satisfies

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

This result tells us how to choose the shift parameter:

$\boxed{\text{Choose } \alpha \text{ just large enough that } \operatorname{Re}\lambda_j + \alpha > 0 \text{ for all } j.}$

"Perturbation" Result

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

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

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 different iterative algorithms:

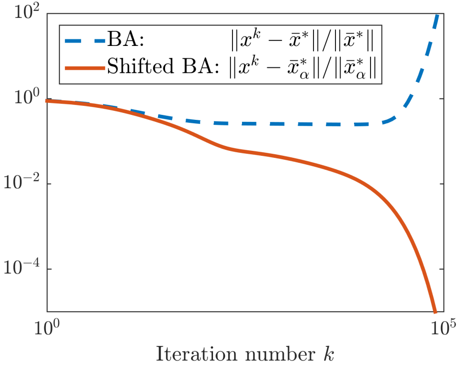
- Matlab's `eigs(_,_,'smallestreal')` (calls ARPACK): baseline algorithm.
- Algorithms by Meerbergen and coauthors: robust but need too many matrix-vector multiplications.
- Krylov-Schur method by Stewart ($\sim$ implicitly restarted Arnoldi): 30% faster than Matlab's `eigs`.
- Jacobi-Davidson: slower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

Our choice: 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.

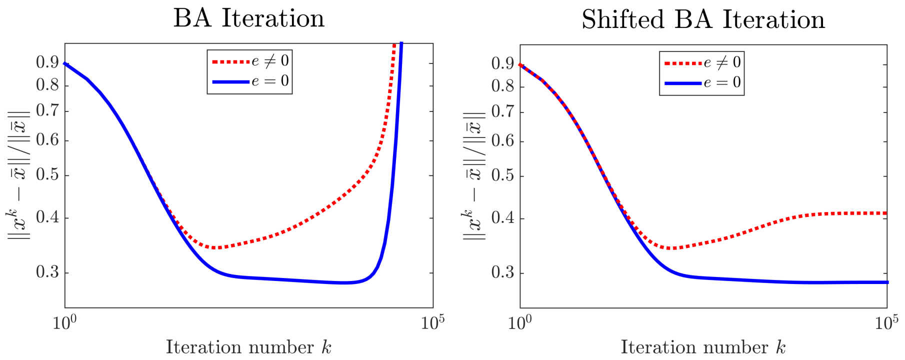
$\rho(BA) = 1.76 \cdot 10^4$
$\alpha = 1.85$



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^*.

Conclusion

- 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.
- Current work focuses on solving the unmatched normal equations $BAx = Bb$ via Krylov subspace methods.