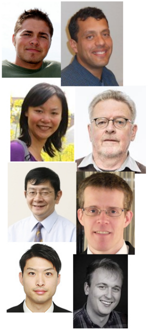
Convergence & Non-Convergence of Algebraic Iterative Reconstruction Methods

Per Christian Hansen

DTU Compute, Technical University of Denmark



Joint work with

Tommy Elfving – Linköping University

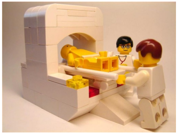
Ken Hayami – NII, Tokyo

Michiel E. Hochstenbach – TU Eindhoven

Keiichi Morikuni – Univ. of Tsukuba

Yiqiu, Nicolai & Jakob – DTU Compute

Emil Y. Sidky – Univ. of Chicago



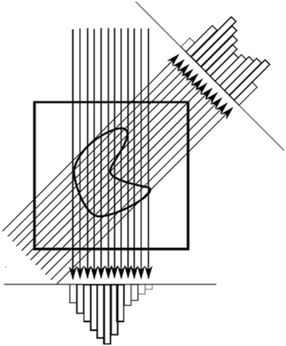
Overview of This Talk

- X-ray CT model
- Stationary iterative reconstruction methods
- Their convergence
- Semi-convergence with noisy data
- Unmatched projectors
- Non-convergence and how to avoid it
- Fixing stationary methods
- Krylov subspace methods $\rightarrow$ GMRES

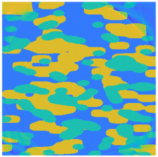
Prelude: X-Ray CT and the Radon Transform

The Principle

Send X-rays through the object f at many angles, and measure the attenuation g.

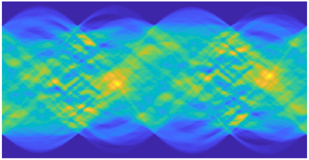


f = 2D object/image



$g = \mathcal{R} f$ = Radon transform of f

= sinogram



$\mathcal{B}$ = adjoint($\mathcal{R}$)

Reconstruction Algorithms

Transform-based methods

- Formulate the forward problem as a certain *transform*, then formulate a stable way to *invert* the transform.
- 2D parallel-beam CT: Radon transform $\leftrightarrow$ filtered back projection.
- Tailored to specific measurement geometries.
- Works well with lots of data.

Algebraic methods

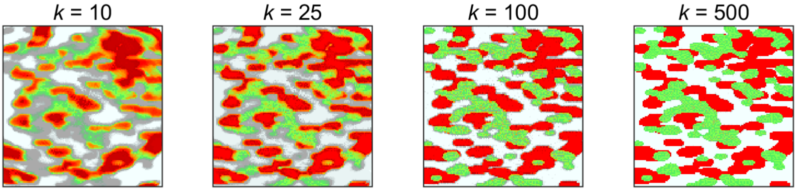
- Discretize the forward problem and solve the corresponding large-scale problem $A\,x = b$ by means of an *iterative method*.
- Works for any measurement geometry.
- Can work well with limited data.

The simplest method: Landweber = steepest descent

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

Example of Convergence for Landweber

Image size: 128×128.
Data: 360 projection angles in $[1^\circ, 360^\circ]$,
181 detector pixels.



We must be concerned with three types of convergence:

1. **Convergence** of the iterative method.
2. **Semi-convergence** in the face of noisy data.
3. **Non-convergence** when forward and back projections don't match.

Asymptotic Convergence for Landweber

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 $\kappa = \|A\|_2 \, \|A^{-1}\|_2$. This is 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 $1 - 2/\kappa^2$.

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:

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

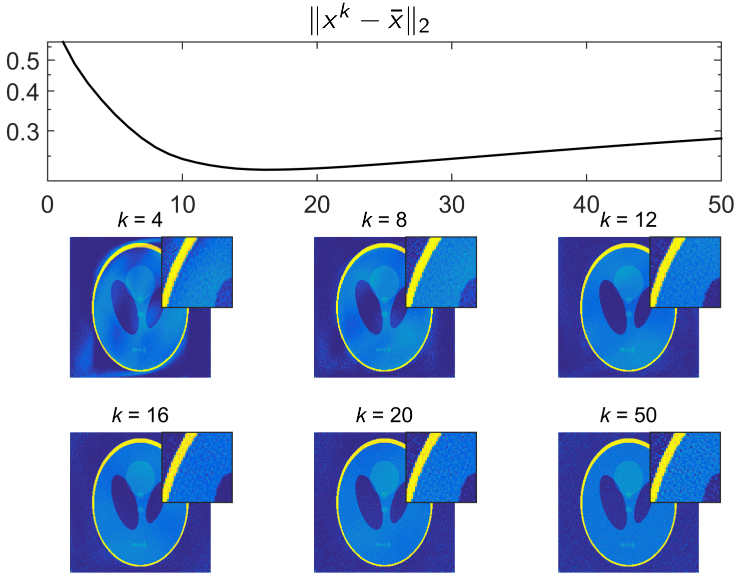
The naïve solution $x^{\text{naïve}} = A^{-1}b$ is undesired, because it has a large component coming from the noise in the data:

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

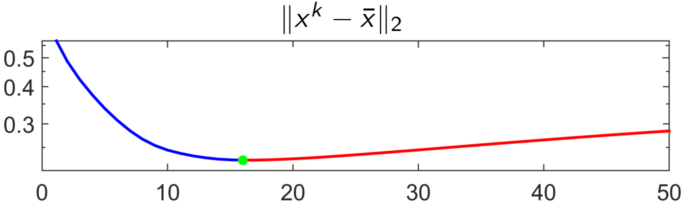
The component $A^{-1}e$ dominates over $\bar{x}$, because A is ill conditioned.

But something interesting happens during the iterations ...

The Reconstruction Error With Noisy Data



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.

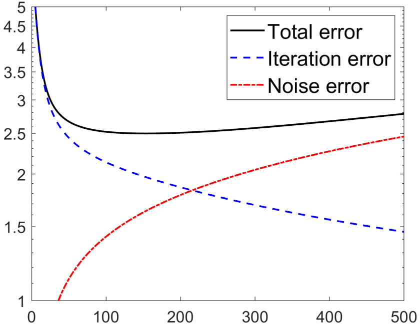
Convergence Analysis: Split the Error

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

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

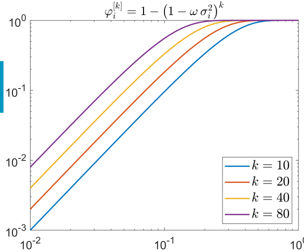
We expect the iteration error to decrease and the noise error to increase.

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



Analysis of Semi-Convergence for Landweber

We use the SVD: $A = \sum_{i=1}^{n} u_i \, \sigma_i \, v_i^T$



Van der Sluis & Van der Vorst (1990)

Filtered SVD solution:

$$x^k = \sum_{i=1}^{n} \varphi_i^{[k]} \, \frac{u_i^T b}{\sigma_i} \, v_i$$

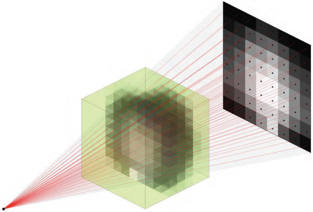
Recall that we solve *noisy* systems $A\,x = b$ with $b = A\,\bar{x} + e$. Then:

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

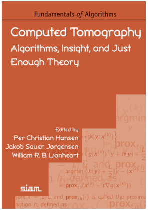
Storage Considerations

$N \times N$ image: each X-ray intersects at most $2N$ pixels $\rightarrow$ at most $2N$ nonzero elements in each row of A (at most $3N$ in 3D) $\rightarrow$ A is *sparse*.

Can still be problematic. 3D example: 1000 projection angles, 1000×1000 detector pixels, $1000 \times 1000 \times 1000$ voxels $\rightarrow$ number of non-zeros in A is of the order 10^{12} $\sim$ several Terabytes of memory.

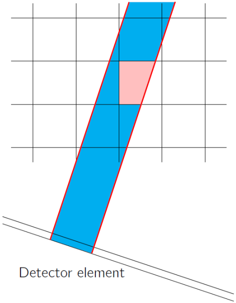


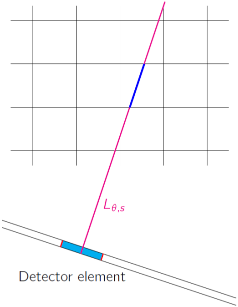
More details here:

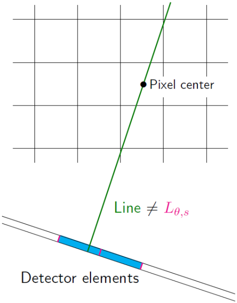


Alternative: use projection models to compute the matrix multiplications – the forward and back projections – "on the fly." We avoid the impossible task of storing A, at the price of having to recompute the matrix elements each time we need them.

Examples of Projection 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

Forward line model: start from detector element centers.

Back projection model: start from image pixel centers and interpolate detector element values.

Projectors and Matrices

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

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

When we can store A then we use A^T for back projection B, and our stationary iterative methods solve least squares problems associated with the normal equations $A^T A\,x = A^T b$.

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

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

We must study the influence of **unmatched** projector pairs on the computed solutions and the convergence of the iterations.

Convergence Analysis for Unmatched Pairs

Substituting B for A^T in Landweber leads to the **BA Iteration**

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

A *fixed-point iteration* that is not related to solving a minimization problem!

Any fixed point x^* satisfies the **unmatched normal equations**

$$\boxed{B\, A\, x^* = B\, b.}$$

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

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

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

Iteration Error and Noise Error When $\mathrm{Re}(\lambda_j) > 0\ \forall j$

Elfving, H (2018)

The *iteration error* is given by

$$\bar{x}^k - \bar{x}^* = T^k(\bar{x}^0 - \bar{x}) , \quad \bar{x}^0 = \text{initial vector} , \quad T = I - \omega\, B\, A ,$$

and it follows that we have **linear** convergence:

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

With $b = A\,\bar{x} + e$ the *noise error* satisfies

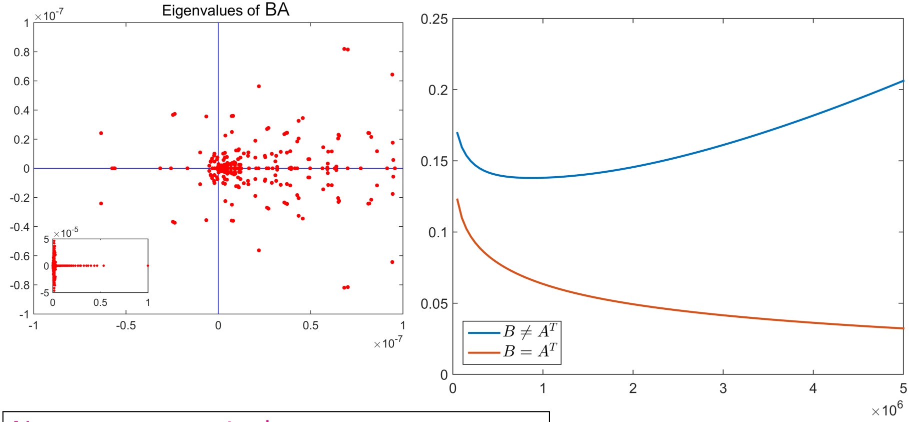
$$\boxed{\|x^k - \bar{x}^k\|_2 \;\leq\; (\omega\, c \|B\|_2)\, k\, \|e\|_2}$$

where we define the constant c by: $\sup_j \|(I - \omega B A)^j\|_2 \leq c$.

I.e., the upper bound grows linearly with the number of iterations k.

Numerical Example of Non-Convergence – **no Noise**

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

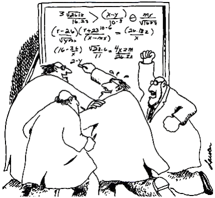


Non-convergence is the most common case.

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.



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.

Many thanks to Tommy Elfving for originally suggesting this.

Convergence Results

Dong, H, Hochstenbach, Riis (2019)

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

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

where $\{\lambda_j\} = \texttt{eig}(BA)$.

The fixed point x^*_α satisfies

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

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

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

Alternative: Solve the Unmatched Normal Equations

Instead of "fixing" a stationary method designed for solving another problem, just solve the unmatched normal equations in one of the forms

UNE: $\boxed{B\,A\,x = B\,b}$ or $\boxed{A\,B\,y = b\,, \quad x = B\,y}$

The left- or right-preconditioned **GMRES** method for (A, b) immediately presents itself as a good choice with B as the preconditioner.

BA-GMRES solves $B\,A\,x = B\,b$ with B as a left preconditioner.

AB-GMRES solves $A\,B\,y = b\,,\ x = B\,y$ with B as a right preconditioner.

Advantages:

- these methods always converge,
- no need for relaxation parameter or shift parameter.

Semi-convergence: Calvetti, Lewis, Reichel (2002), Gazzola, Novati (2016).

Solving the Unmatched Normal Equations

Hayami, Yin, Ito (2010)

AB-GMRES solves $\min_y \|A\,B\,y - b\|_2$, $x = B\,y$ $\quad$ (B = right precond.)

- $\min_x \|A\,x - b\|_2 = \min_z \|A\,B\,z - b\|_2$ holds for all b if an only if $\text{range}(B) = \text{range}(A^T)$.
- The LS residual norm $\|A\,x^k - b\|_2$ decreases monotonically.

BA-GMRES solves $\min_x \|B\,A\,x - B\,b\|_2$ $\quad$ (B = left preconditioner)

- the problems $\min_x \|A\,x - b\|_2$ and $\min_x \|B\,A\,x - B\,b\|_2$ are equivalent for all b if and only if $\text{range}(B^T) = \text{range}(A)$.
- The UNE residual norm $\|B\,A\,x^k - B\,b\|_2$ decreases monotonically.

Both methods use the same Krylov subspace

$$\text{span}\{B\,b, B\,A\,B\,b, \ldots, (B\,A)^{k-1}B\,b\}$$

for the solution, but they use different objective functions.

When the Transpose Matches

H, Hayami, Morikuni

AB-GMRES with $B = A^T$ computes $x^k = A^T u^k$ with

$$u_k = \arg \min_{u \in \mathcal{K}_k(A A^T, b)} \|A A^T u - b\|_2^2 \ .$$

The method minimizes $\|b - A x^k\|_2$ and so do **CGLS** and **LSQR**; they produce the same iterates (in ∞ precision).

BA-GMRES with $B = A^T$ applies GMRES to

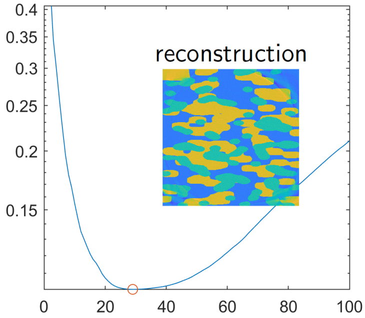
$$A^T A x = A^T b \qquad \longleftrightarrow \qquad \min_x \|A x - b\|_2 \ .$$

Equivalent to applying MINRES to the normal equations $A^T A x = A^T b$ which, in turn, is equivalent to **LSMR** (in ∞ precision).

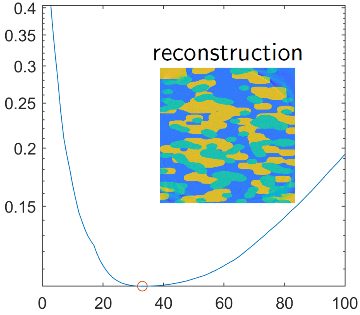
Reconstr. Error, Noisy Data, Matrix is 252 000 × 176 400

Image has 420 × 420 pixels, 600 projection angles, 420 detector pixels.

AB-GMRES $\|x^k - \bar{x}\|_2/\|\bar{x}\|_2$



BA-GMRES $\|x^k - \bar{x}\|_2/\|\bar{x}\|_2$



- Semi-convergence is evident.
- Same minimum reconstruction error $\|x^k - \bar{x}\|_2/\|\bar{x}\|_2 \approx 0.10$ for both.
- Slightly fewer iterations for AB-GMRES in this example.

Stopping Rules

We must terminate the iterations at the point of semi-convergence.

- **Discrepancy principle (DP):** terminates the iterations as soon as the residual norm is smaller than the noise level:

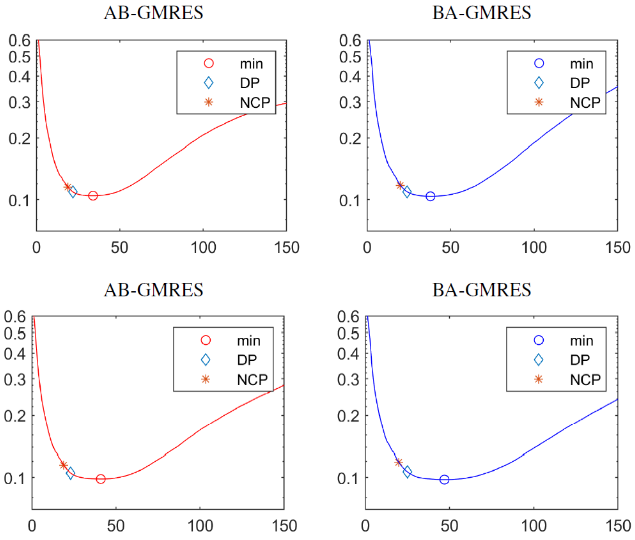
  $$k_{\mathrm{DP}} = \text{the smallest } k \text{ for which } \|b - A\,x^k\|_2 \le \tau\,\|e\|_2$$

  where $\tau \ge 1 =$ safety factor when we have a rough estimate of $\|e\|_2$.

- **NCP criterion:** uses a normalized cumulative periodogram to perform a spectral analysis of the residual vector $b - A\,x^k$ and identifies when the residual is close to being white noise – indicating that all available information has been extracted from the noisy data.

For those who are curious: the L-curve criterion does not work, and we cannot implement generalized cross validation (GCV) efficiently.

Stopping Rules: Tests With 2 Different Back Projectors



Both DP and NCP stop a bit too early – better than stopping too late.

Brand New Results[1]

Iterative image reconstruction for CT with unmatched projection matrices using the generalized minimal residual algorithm

Emil Y. Sidky, Per Christian Hansen, Jakob S. Jørgensen, and Xiaochuan Pan

Cone-beam CT data acquired on an Epica Pegaso veterinary CT scanner with 180 projections taken uniformly over one circular rotation, using a physical "quality assurance phantom."

Detector: 1088×896 pixels. Reconstruction: $1024 \times 1024 \times 300$ voxels.

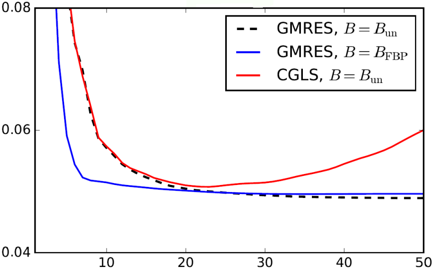
Ray-driven forward projector A. Two choices of back projector B:

- B_{un} = voxel-driven back projection, linear interpolation on detector.
- $B_{\mathrm{FBP}} = B_{\mathrm{un}} F$ = filtered back-projection, where F = ramp filter.

[1] 7th Intl Conf on Image Formation in X-Ray CT, June 12-16, 2022, Baltimore.

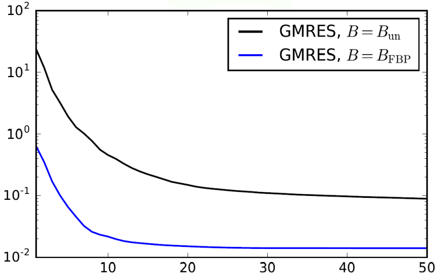
Brand New Results – Convergence

$\|Ax^k - b\|_2$



Decay: no guarantee

$\|BAx^k - Bb\|_2$



Decay: guaranteed

BA-GMRES works well with both B-matrices.

“CGLS” fails to converge, because it is not CGLS.

Coda

Convergence

- Good understanding of convergence for noise-free data.
- Emerging: good understanding of semi-convergence for noisy data.
- Non-convergence of stationary methods due to unmatched projectors.
- Can be fixed $\rightarrow$ shifted BA iteration.
- Not an issue for AB-GMRES and BA-GMRES.

Method	$+$	$-$
Shifted BA	no extra storage	2 parameters ω and α
ABBA GMRES	no extra parameters	storage for Arnoldi vectors

Future

- More theory about semi-convergence for GMRES.
- Deal with GMRES memory issue: restart, recycling, etc.
- Ready-to-use implementations for the CT community.