GMRES-Based Methods for Unmatched Projectors in X-Ray CT

Per Christian Hansen

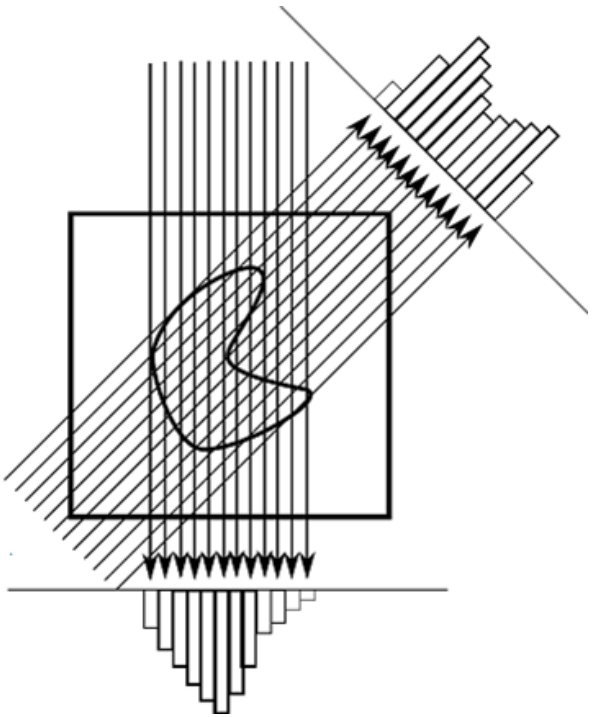
Technical University of Denmark

Joint work with
Ken Hayami, NII, Tokyo
Keiichi Morikuni, Univ. Tsukuba
Emil Sidky, Univ. Chicago

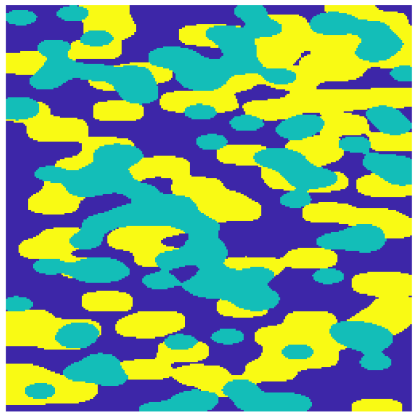
Prologue: X-Ray CT in 2D – and the Radon Transform

The Principle

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

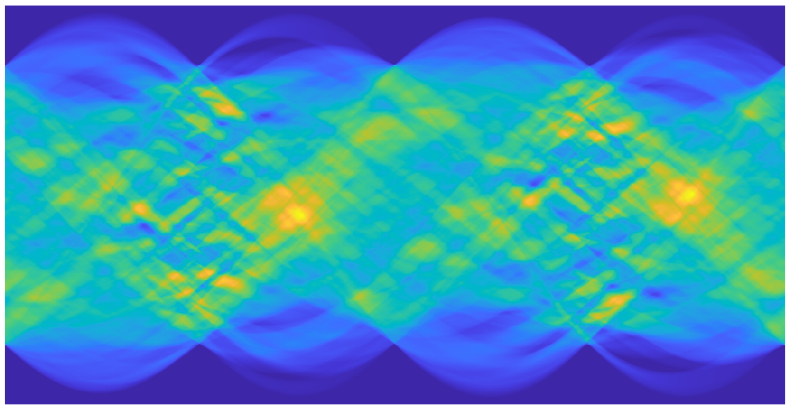


f = 2D object/image



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

= sinogram



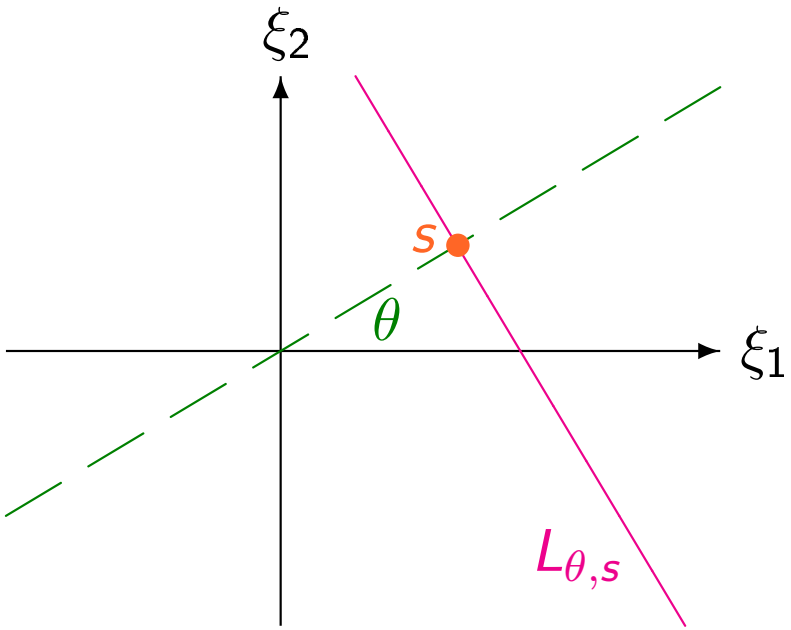
Prologue: Forward and Back Projections

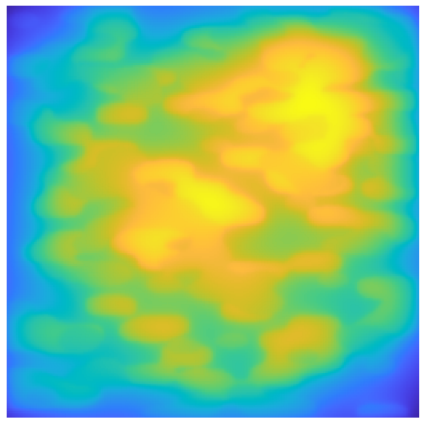
Forward projection $\mathcal{R}$, the Radon transform models the scanner physics via integration of the function f along lines $L_{\theta,s}$

$$\mathcal{R}[f](\theta, s) = \int_{L_{\theta,s}} f(\xi_1, \xi_2)\, d\ell = g(\theta, s) = \text{sinogram} \ .$$

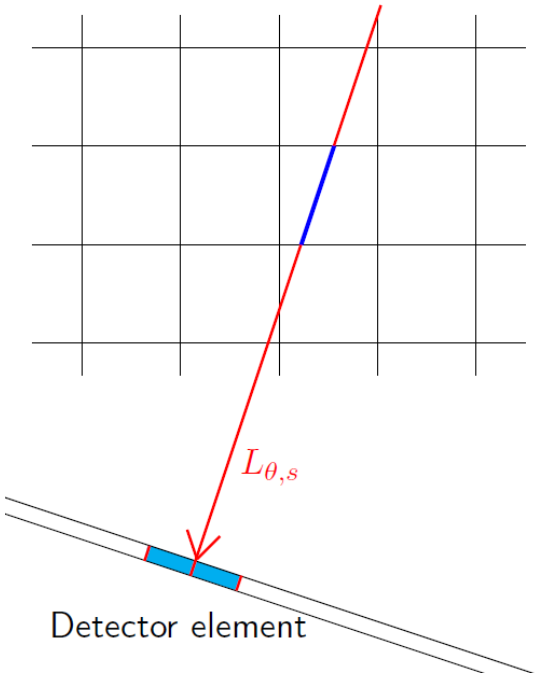
Back projection $\mathcal{B} = \text{adjoint}(\mathcal{R})$, an abstraction, smears g back along $L_{\theta,s}$

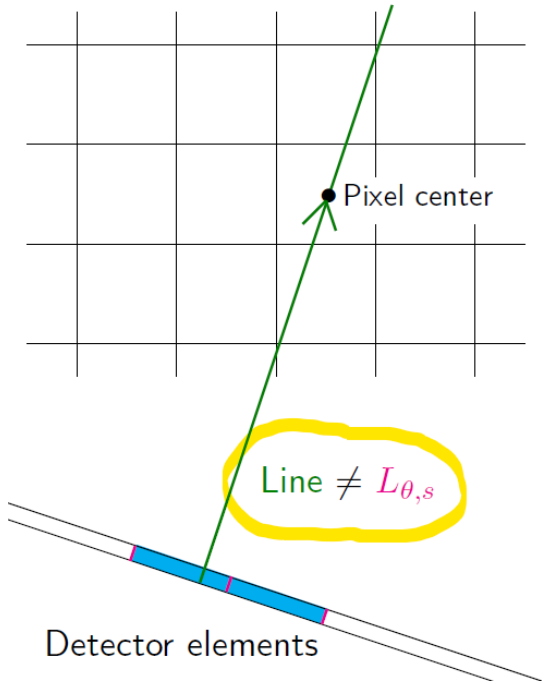
$$\mathcal{B}[g](\xi_1, \xi_2) = \int_0^{2\pi} g(\theta, \xi_1 \cos\theta + \xi_2 \sin\theta)\, d\theta \ .$$

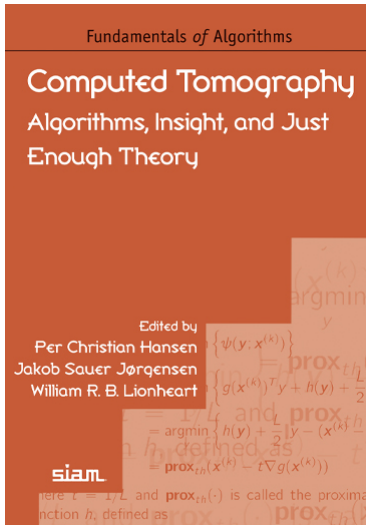




Ray/Pixel Driven Discretization Models







Forward line model

Ray driven

Back projection model

Pixel driven

Details here

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

When we can store A (it is *sparse*), then we can use $B = A^T$, and solve the normal equations $A^T A\, x = A^T b$ associated with the least squares problem, with x = vec(image) and b = vec(sinogram).

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

Ray and pixel driven models $\rightarrow$ $B \neq A^T$ $\rightarrow$ unmatched projector pair.

An unmatched transpose also arises in image deblurring with anti-reflective boundary conditions, due to different boundary conditions for the blurring and its adjoint (Donatelli et al. 2006). It is handled by the same approach as used in this work (Donatelly et al. 2015).

Solve the Unmatched Normal Equations

Classical iterative methods (e.g., Cimmino, SIRT, CGLS) require $B = A^T$.
Alternatively, we can solve the **unmatched normal equations**

$$\textbf{UNE:} \quad \boxed{BAx = Bb} \qquad \text{or} \qquad \boxed{ABy = b\,, \quad x = By}$$

We will use **GMRES** (Saad, Schultz, 1986), a very efficient iterative method for solving systems

$$\boxed{Mx = d} \qquad \text{with a square and nonsymmetric matrix } M.$$

We skip the implementation details here, and just remind that in the kth step, the iterate x^k of GMRES solves the problem

$$\min_x \|Mx - d\|_2 \qquad \text{subject to} \qquad x \in \mathcal{K}_k(M, d)\,,$$

with the *Krylov subspace*

$$\mathcal{K}_k(M, d) = \text{span}\{d,\, Md,\, M^2 d,\, \ldots,\, M^{k-1} d\}\,.$$

AB–GMRES and BA–GMRES

We can formulate *specialized versions* of GMRES for the UNEs:

BA–GMRES solves $BAx = Bb$.

AB-GMRES solves $ABy = b,\ x = By$.

Both methods use the same Krylov subspace $\mathcal{K}_k(BA, Bb)$ for the solution, but they use different objective functions.

Advantages:

- both methods always converge in the absence of noise,
- no need for relaxation parameter (as in Cimmino etc.),
- fairly simple to implement → next page.

▷ K. Hayami, J.-F. Yin, T. Ito, *GMRES methods for least squares problems*, SIAM J. Matrix Anal. Appl., 31 (2010), 2400–2430.

▷ H, K. Hayami, K. Morikuni, *GMRES methods for tomographic reconstruction with an unmatched back projector*, J. Comp. Appl. Math., 413 (2022), 114352.

The ABBA Algorithms

Algorithm AB-GMRES

Choose initial x_0
$r_0 = b - A x_0$
$w_1 = r_0 / \|r_0\|_2$
for $k = 1, 2, \ldots$
  $q_k = AB\, w_k$
  for $i = 1, 2, \ldots, k$
    $h_{i,k} = q_k^{\mathsf{T}} w_i$
    $q_k = q_k - h_{i,k}\, w_i$
  endfor
  $h_{k+1,k} = \|q_k\|_2$
  $w_{k+1} = q_k / h_{k+1,k}$
  $y_k = \arg\min_y \| \, \|r_0\|_2 \, e_1 - H_k \, y \, \|_2$
  $x_k = x_0 + B\,[w_1, w_2, \ldots, w_k]\, y_k$
  $r_k = b - A\, x_k$
  stopping rule goes here
endfor

Algorithm BA-GMRES

Choose initial x_0
$r_0 = B\, b - B\, A\, x_0$
$w_1 = r_0 / \|r_0\|_2$
for $k = 1, 2, \ldots$
  $q_k = B\, A\, w_k$
  for $i = 1, 2, \ldots, k$
    $h_{i,k} = q_k^{\mathsf{T}} w_i$
    $q_k = q_k - h_{i,k}\, w_i$
  endfor
  $h_{k+1,k} = \|q_k\|_2$
  $w_{k+1} = q_k / h_{k+1,k}$
  $y_k = \arg\min_y \| \, \|r_0\|_2 \, e_1 - H_k \, y \, \|_2$
  $x_k = x_0 + [w_1, w_2, \ldots, w_k]\, y_k$
  $r_k = b - A\, x_k$
  stopping rule goes here
endfor

MATLAB and Python software available from PCH at
https://people.compute.dtu.dk/pcha/ABBA
and in TIGRE: Tomographic Iterative GPU-based Reconstruction Toolbox
https://github.com/CERN/TIGRE

Obligatory Slide with Many Equations

Hayami, Yin, Ito (2010), H, Hayami, Morikuni (2022)

AB–GMRES solves $\min_y \|A\,B\,y - b\|_2$, $x = B\,y$

- $\min_x \|A\,x - b\|_2 = \min_y \|A\,B\,y - b\|_2$ holds for all b if an only if $\text{range}(A\,B) = \text{range}(A)$, e.g., if $\text{range}(B) = \text{range}(A^T)$.
- Monotonic decay of $\|A\,x^k - b\|_2$.
- Equivalent to LSQR when $B = A^T$.

BA–GMRES solves $\min_x \|B\,A\,x - B\,b\|_2$

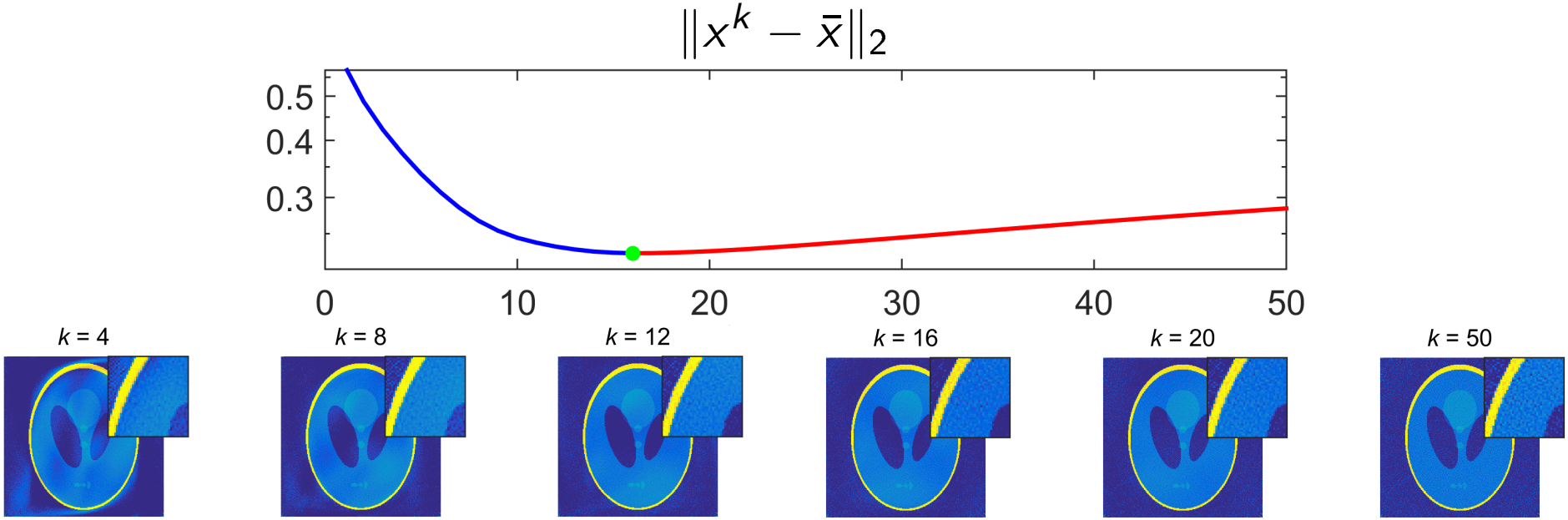
- the problems $\min_x \|A\,x - b\|_2$ and $\min_x \|B\,A\,x - B\,b\|_2$ are equivalent for all b if and only if $\text{range}(B^T B\,A) = \text{range}(A)$, e.g., if $\text{range}(B^T) = \text{range}(A)$.
- Monotonic decay of $\|B\,A\,x^k - B\,b\|_2$.
- Equivalent to LSMR when $B = A^T$.

Conditions are difficult/impossible to check in a given CT problem.

Iterative Methods: Noisy Data Gives Semi-Convergence

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:

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

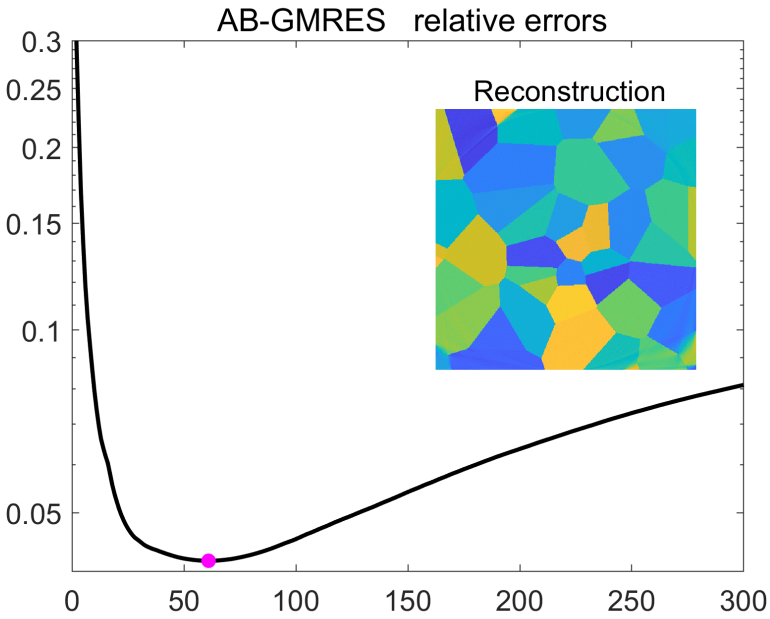


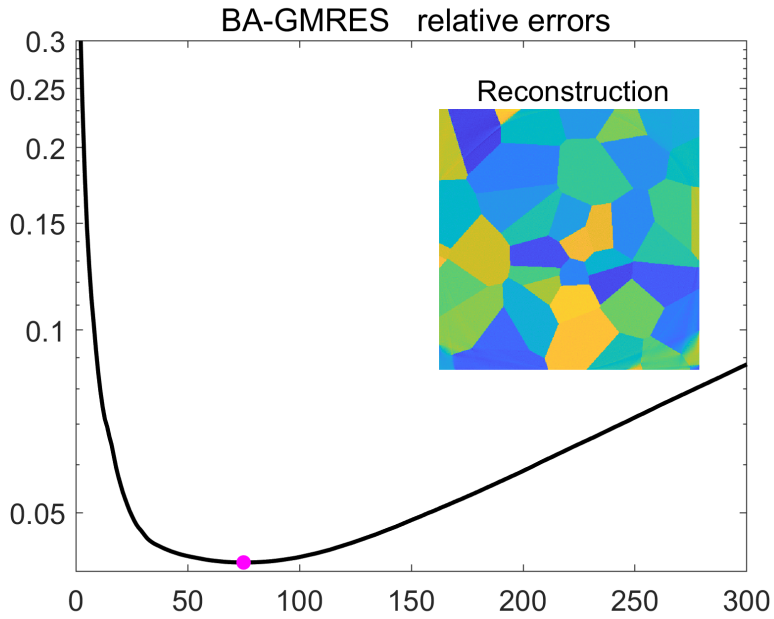
- In the initial iterations x^k approaches the unknown ground truth $\bar{x}$.
- During later iterations x^k converges to the undesired $x^{\text{naïve}} = A^{-1}b$.
- Stop the iterations when the convergence behavior changes.

ABBA Reconstruction Errors $\|x^k - \bar{x}\|_2 / \|\bar{x}\|_2$

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

A and B generated with GPU-ASTRA software; A is $252\,000 \times 176\,400$.



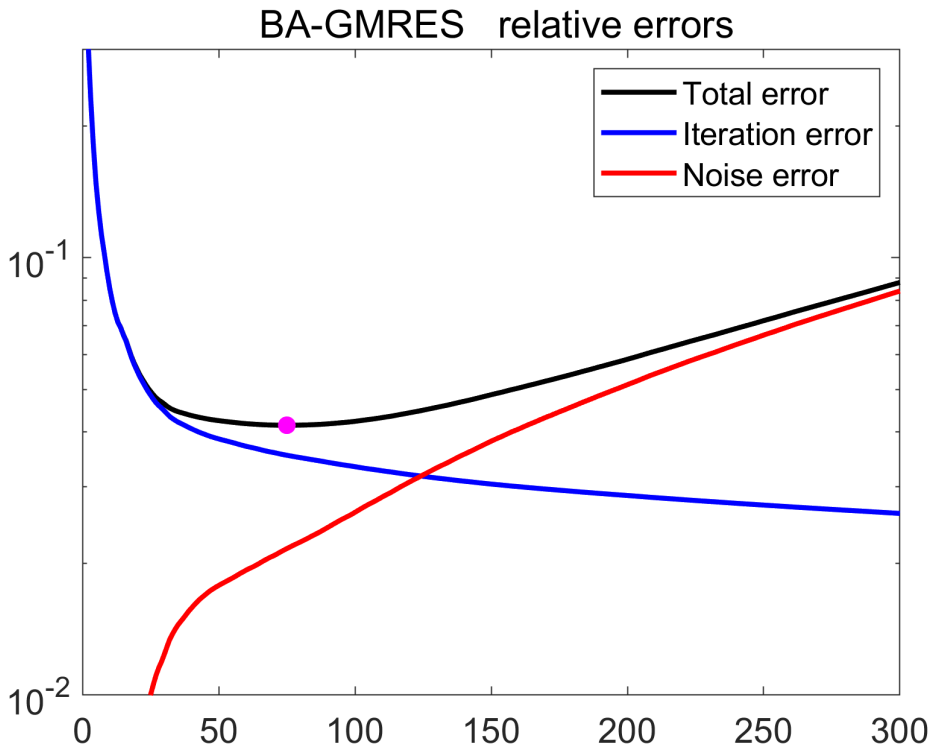


- Semi-convergence is evident for both methods.
- Same minimum reconstruction error $\|x^k - \bar{x}\|_2 / \|\bar{x}\|_2 \approx 0.042$ for both.
- Slightly fewer iterations for AB–GMRES in this example.
- Storage for Krylov bases – AB–GMRES: $m \times k$ – BA–GMRES: $n \times k$.

Error Propagation in BA-GMRES

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

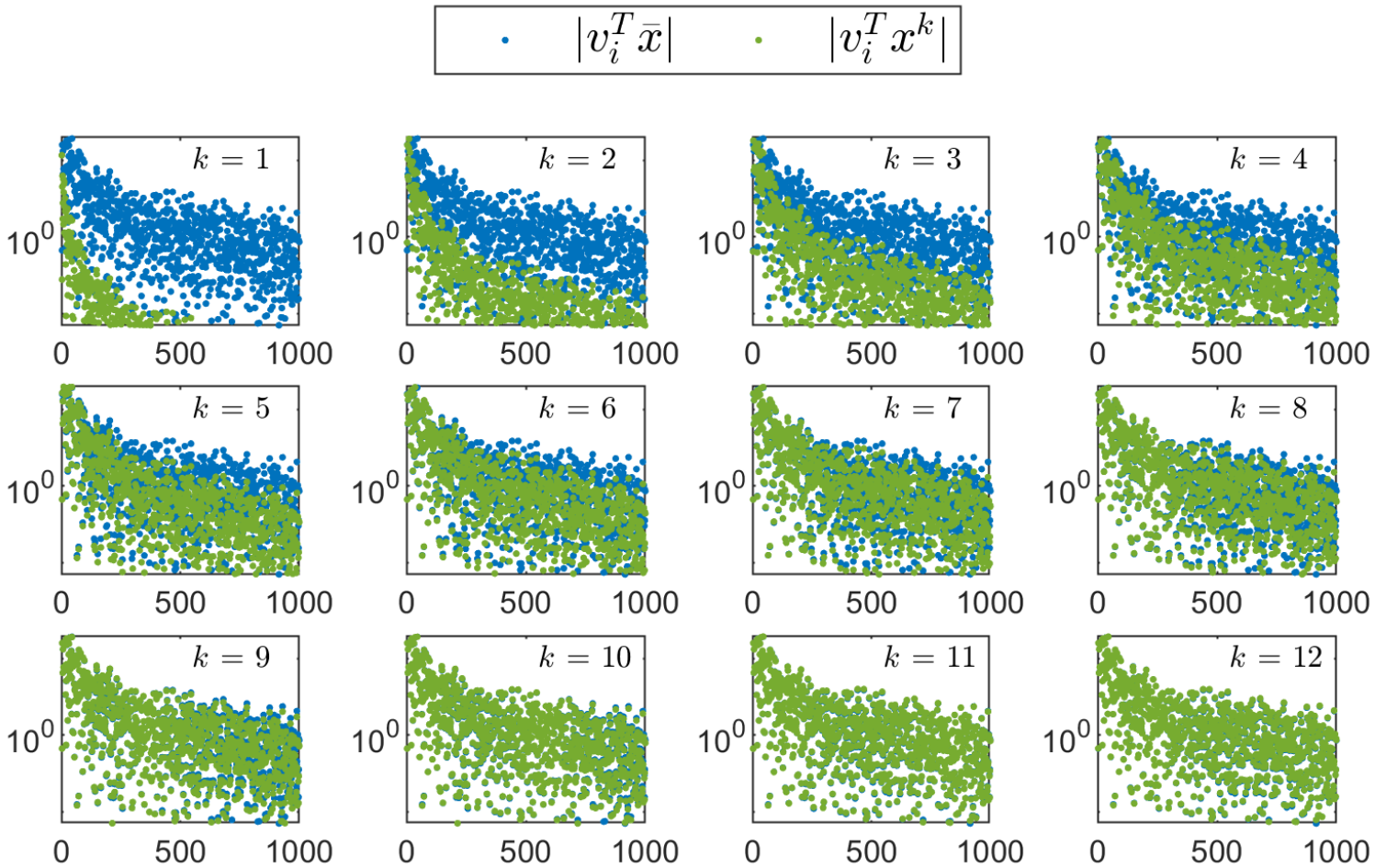
$$\underbrace{x^k - \bar{x}}_{\text{total error}} = \underbrace{x^k - \bar{x}^k}_{\text{noise error}} + \underbrace{\bar{x}^k - \bar{x}}_{\text{iteration error}}$$



Both errors are monotonic in this example – but no guarantee for this.

Semi-Convergence of BA–GMRES ★ SVD Insight

Recall: $b = A\bar{x} + e$, $\bar{x}$ = ground truth, $\|e\|_2/\|\bar{b}\|_2 = 0.001$ Gaussian.



- As k increases we capture more and more SVD components in x^k.
- At $k = 11$ we already capture the first 1000 SVD components.

Semi-Convergence of BA–GMRES ★ Polynomial Insight

BA–GMRES is a Krylov subspace method, with Krylov subspace

$$\mathcal{K}_k(BA, Bb) = \text{span}\{Bb, BA\,Bb, \ldots, (BA)^{k-1}\,Bb\}$$

and with associated polynomial $\mathcal{P}_k$, which is constructed when applying BA–GMRES to the noisy right-hand side b.

We can write the BA–GMRES iterates as

$$x^k = \mathcal{P}_k(BA)\,Bb = \mathcal{P}_k(BA)\,B\bar{b} + \mathcal{P}_k(BA)\,Be\ ,$$

where the latter term is the noise error.

How to determine the roots of $\mathcal{P}_k$ is described in

S. Goossens & D. Roose, *Ritz and harmonic Ritz values and the convergence of FOM and GMRES,* NLAA, 6 (1999), 281–293.

Sensitive to rounding errors – don't know a stable approach (yet).

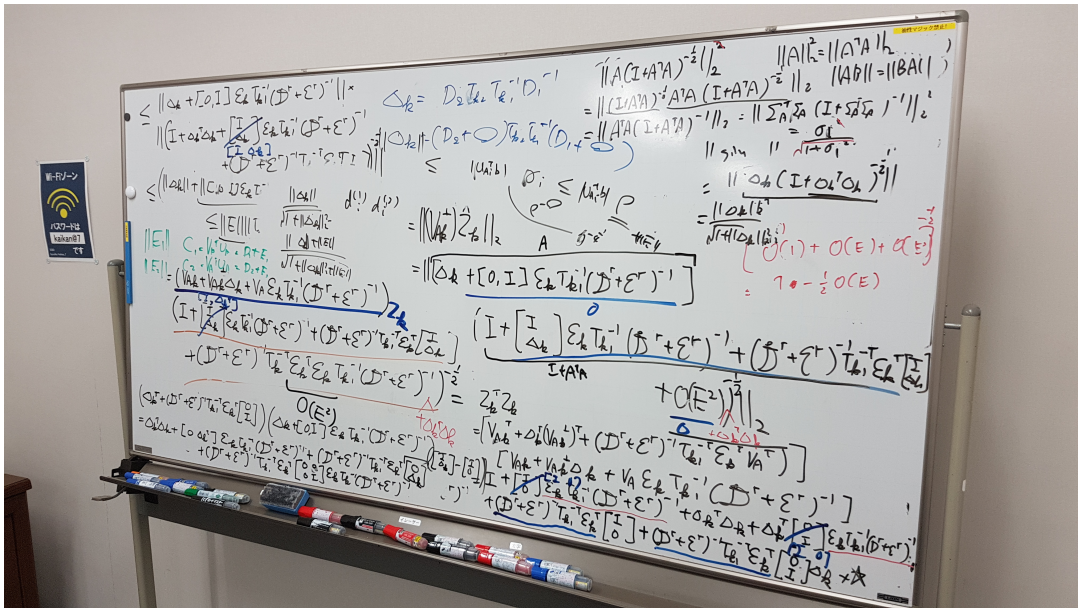
Semi-Convergence of BA–GMRES ★ Subspace Insight

In our numerical example, $\|B - A^T\|_F / \|B\|_F \approx 0.15$.

So how come we can still compute good reconstructions?

- Compare the Krylov subspace with the SVD subspace.

We are working on it ...



PCH and Keiichi Morikuni in snowy Tokyo

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:

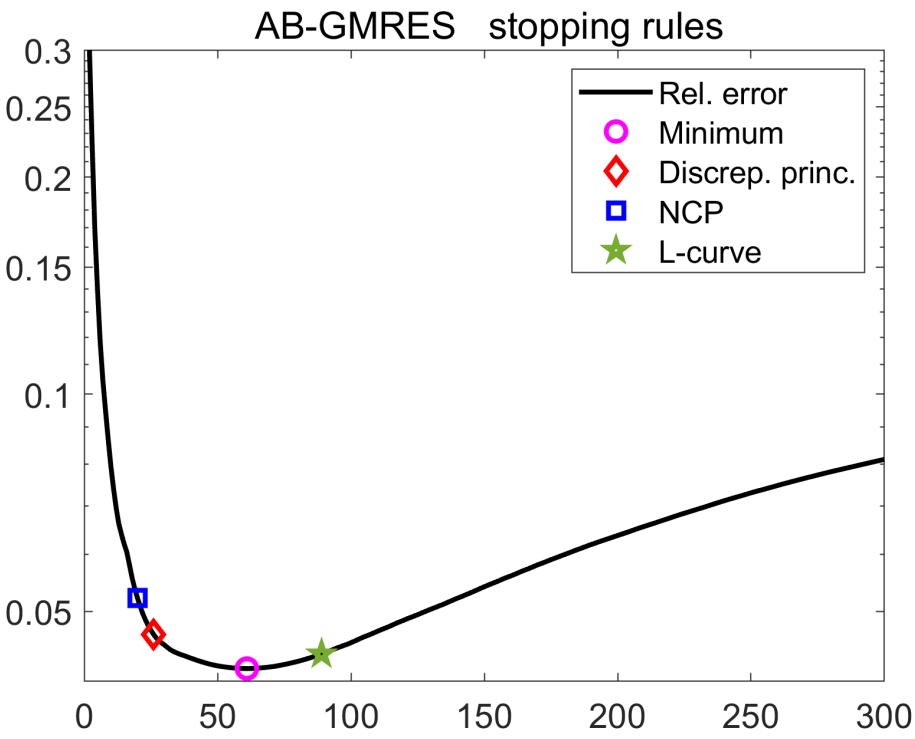
$$k_{\mathrm{DP}} = \text{the smallest } k \text{ for which } \|b - A x^k\|_2 \leq \tau \, \|e\|_2$$

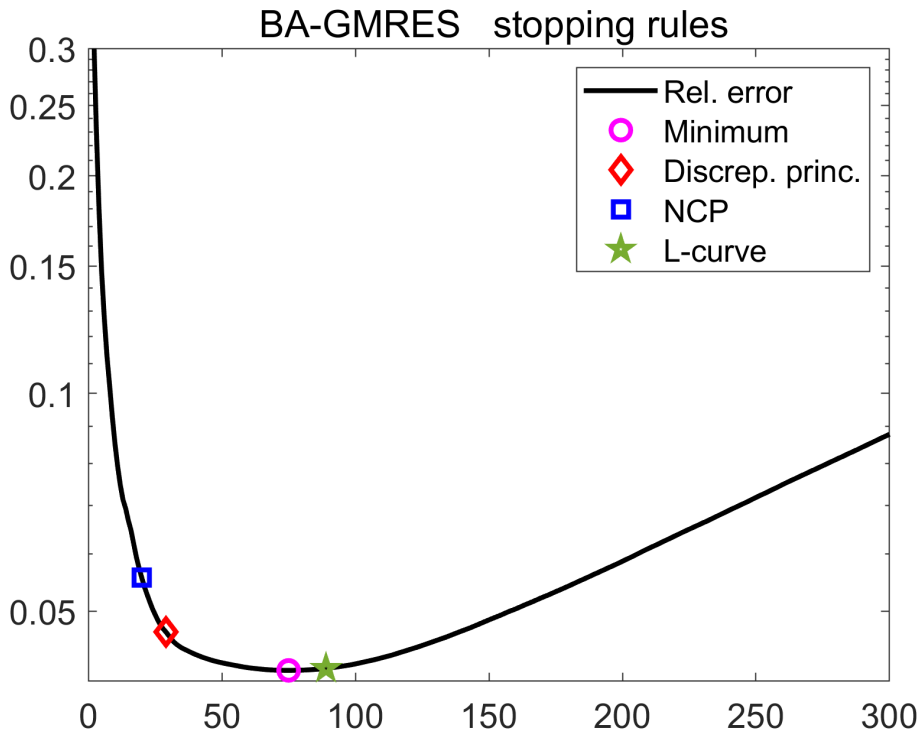
  where $\tau \geq 1$ = safety factor when we have a rough estimate of $\|e\|_2$.

- **NCP criterion:** uses the 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 – which indicates that all available information has been extracted from the noisy data.

- **L-curve criterion:** locates the "corner" of the L-shaped point set $\big(\log \|b - A x^k\|_2 \, , \, \log \|x^k\|_2 \big)$.

Stopping Rules in Action





Both the discrepancy principle and the NCP criterion stop too early.

The L-curve criterion stops too late.

A topic for further research, related to all iterative regularization methods.

7th International Conference on Image Formation in X-Ray Computed Tomography, June 12–16, 2022, Baltimore, USA

Computational experiments with real data by:

Emil Y. Sidky, Department of Radiology, University of Chicago.

- Cone-beam CT data from an Epica Pegaso veterinary CT scanner.
- 180 projections taken uniformly over one circular rotation.
- Physical "quality assurance" (QA) phantom $\longrightarrow$

- Detector: 1088×896 pixels of size $(0.278\text{mm})^2$.
- 3D reconstruction: $1024 \times 1024 \times 300$ voxel grid.

Ray-driven projector A. Two choices of B:

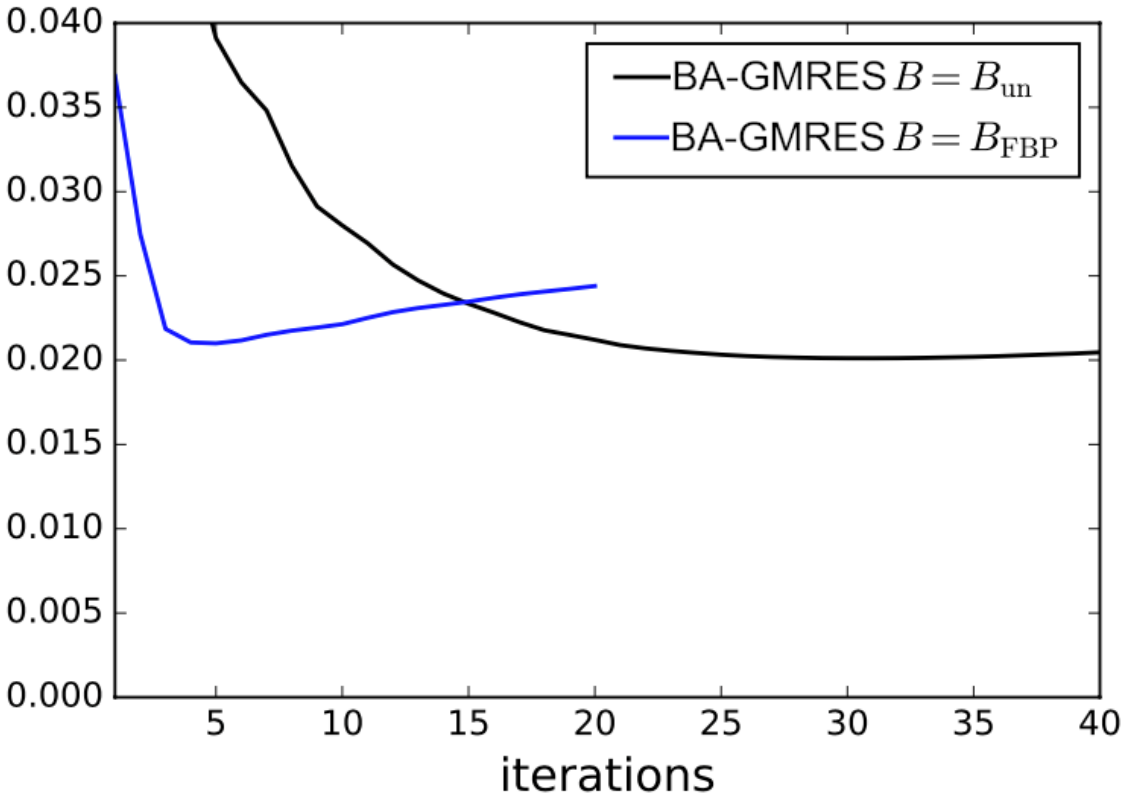
- B_{un} = voxel-driven back projection, linear interpolation on detector.
- $B_{\text{FBP}} = B_{\text{un}} F$ = filtered back-projection, where F = ramp filter.

"Reconstruction Error"

Real data from a physical phantom $\Rightarrow$ no ground truth $\bar{x}$.

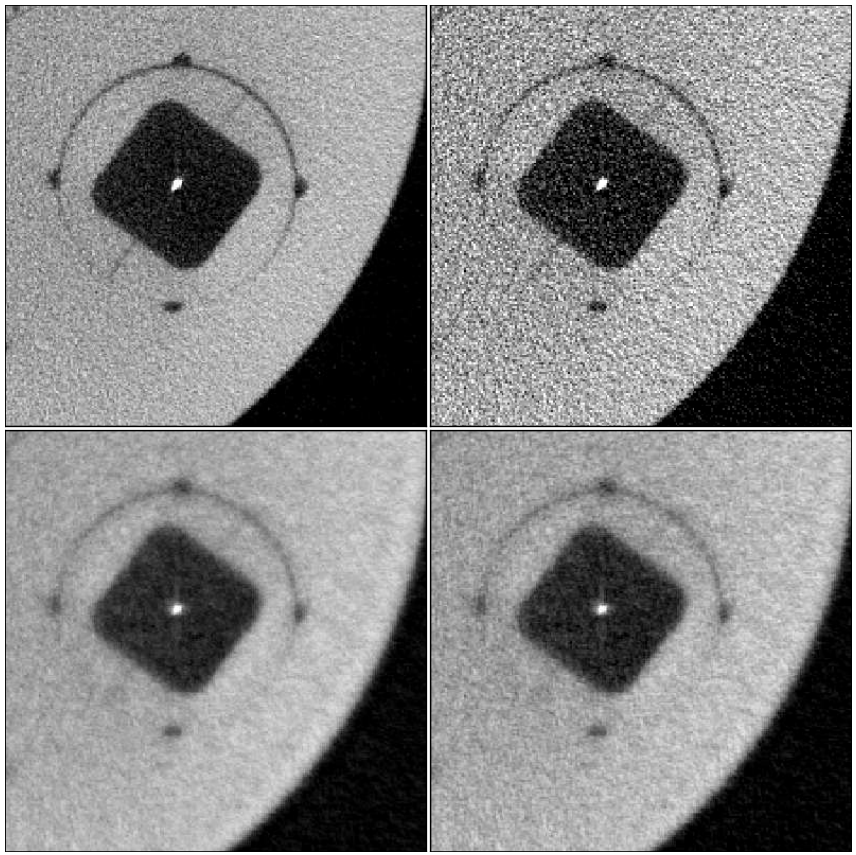
Instead we use a high-quality FBP reconstruction x_{FBP}.

$$\|x^k - x_{\mathrm{FBP}}\|_2$$



With both B-matrices we observe semi-convergence.

Reconstruction, Mid-Slice Region of Interest

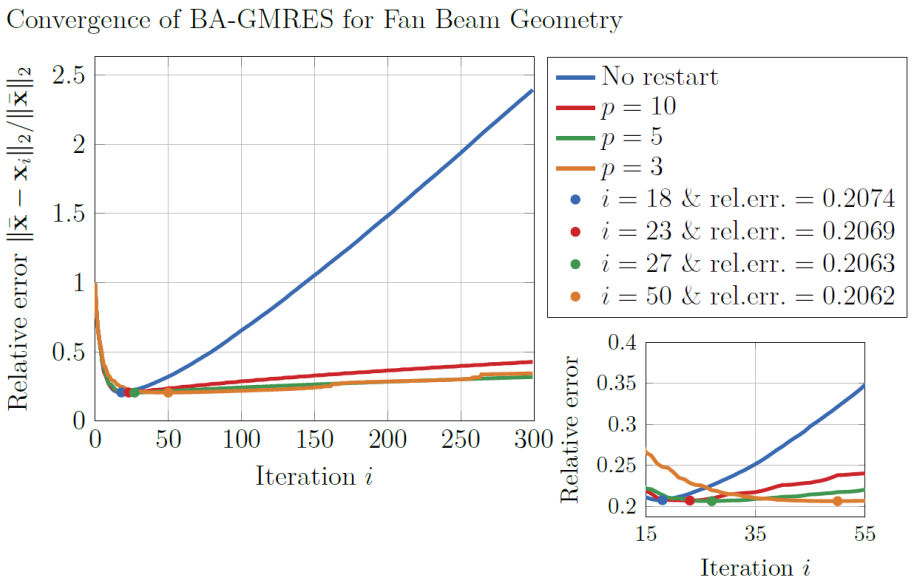


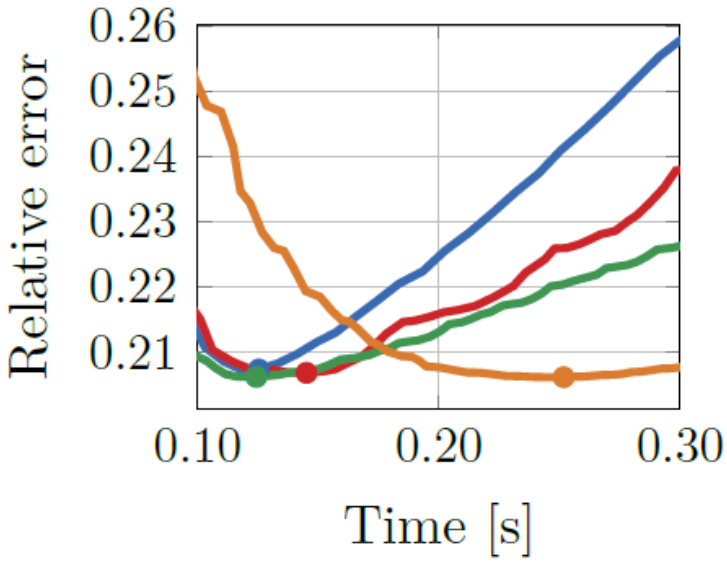
Top left: reference x_{FBP} = FBP reconstructed image from 720 views.
Top right: FBP reconstructed images from 180 views.
Bottom left: BA–GMRES image, $B = B_{\mathsf{un}}$, $k = 29$ iterations, 180 views.
Bottom right: BA–GMRES image, $B = B_{\mathsf{FBP}}$, $k = 4$ iterations, 180 views.

Epilogue: Restart of the ABBA Methods

Must compute and store the orthonormal basis for the Krylov subspace. But orthonormalization is time consuming, and storage may be prohibitive.

Restart solves these problems; more iterations are necessary, but the computing time does not deteriorate.





M. Knudsen, *ABBA Iterative Methods for X-Ray Computed Tomography*, MSc Thesis, DTU, 2023 link

Conclusion

Unmatched projector pairs★

- Need efficient iterative reconstruction methods for unmatched pairs.
- Modify a classical method, e.g., as in the Shifted BA Iteration.
- Use a method that solves the unmatched normal equations → ABBA.

★New matched pair: K. Bredies & R. Huber, *Convergence analysis of pixel-driven Radon and fanbeam transforms*, SIAM J. Numer. Anal., 59 (2021), 1399–1432.

Convergence

- Good understanding of convergence for noise-free data.
- Emerging: understanding of semi-convergence for noisy data.

Future

- More theory about semi-convergence for GMRES.

Appendix: Semi-Convergence ★ Technical Insight

Semi-convergence of CGLS is well understood; but only few results have been obtained for GMRES.

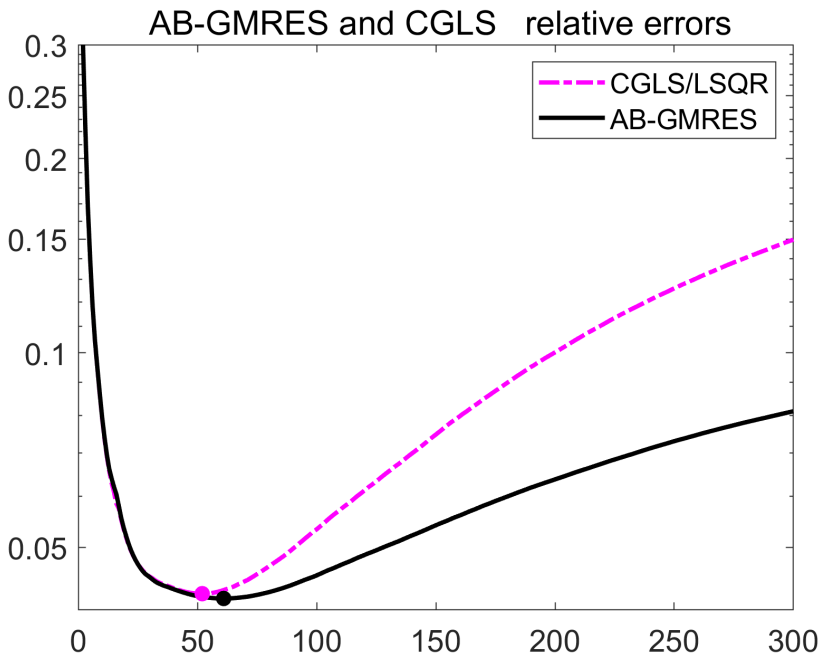
- **Calvetti, Lewis, Reichel (2002):** if the noise-free data lies in a finite-dimensional Krylov subspace, and if GMRES is equipped with a suitable stopping rule, then the GMRES-solution tends towards the exact solution $\bar{x}$ as the noise goes to zero.
- **Gazzola, Novati (2016):** if the discrete Picard condition (DPC) is satisfied and if the left singular vectors of the Hessenberg matrices of two consecutive GMRES steps resemble each other – then the Hessenberg systems in GMRES also satisfy the DPC.

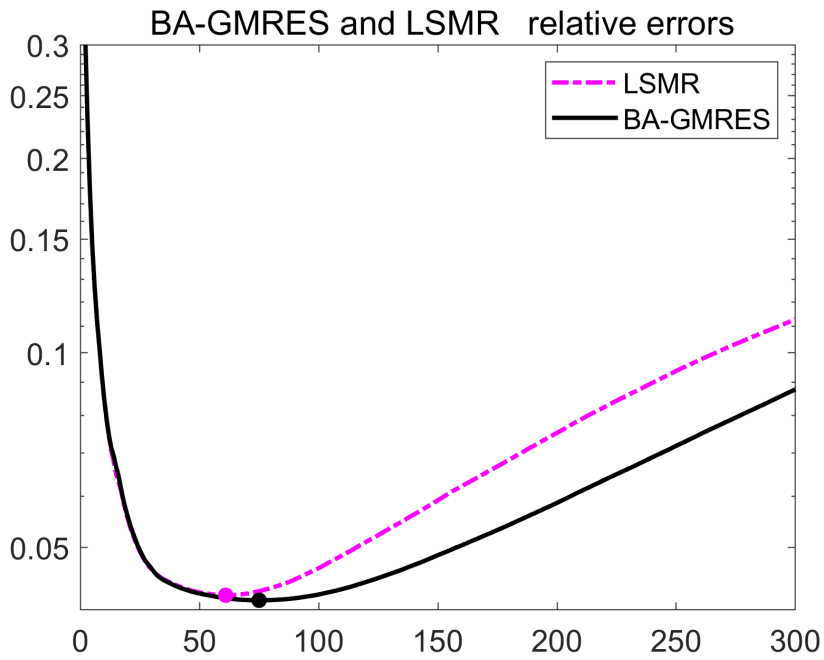
The difficulty is that we cannot analyze GMRES by means of the SVD of A.

A complete understanding has not emerged yet. Here we rely on some preliminary analysis and insight from numerical experiments.

Appendix: Comparison with CGLS/LSQR and LSMR

If $B = A^T$ then AB–GMRES = CGLS/LSQR and BA–GMRES = LSMR.





Recall: for large-scale problems we do not have a choice; we must use B.

The good news is that the reconstruction error does not deteriorate, compared with using A^T (in this example, the error is slightly smaller).