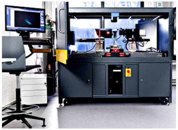
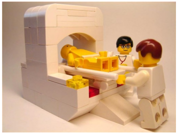
GMRES Methods for Tomographic Reconstruction with an Unmatched Back Projector

Per Christian Hansen
Technical University of Denmark

Joint work with Ken Hayami and Keiichi Morikuni funded by JSPS

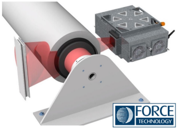


Lab scanner



Medical scanner

Synchrotron

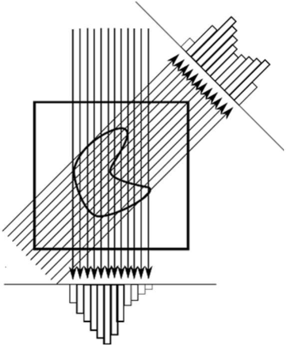


Pipe inspection

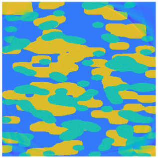
X-Ray CT in 2D – and the Radon Transform

The Principle

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

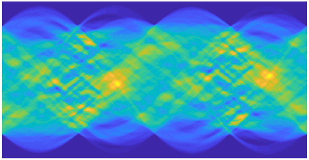


f = 2D object/image



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

= sinogram



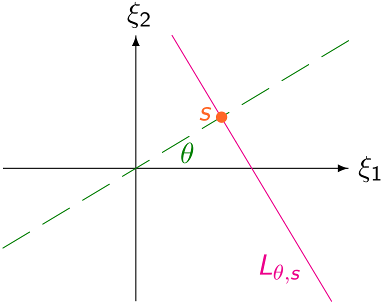
Modelling in CT: Forward and Back Projections

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

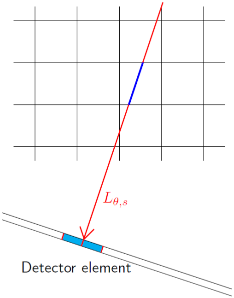
$$\mathcal{R}[f](\theta,s) = \int_{L_{\theta,s}} f(\xi_1,\xi_2)\, d\ell = g(\theta,s) = \text{sinogram} \ .$$

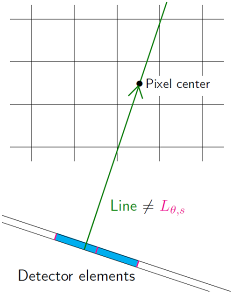
Back projection $\mathcal{B} = \text{adjoint}(\mathcal{R})$, an abstraction, smears g back along $L_{\theta,s}$

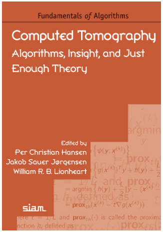
$$\mathcal{B}[g](\xi_1,\xi_2) = \int_0^{2\pi} g(\theta, \xi_1 \cos\theta + \xi_2 \sin\theta)\, d\theta \ .$$



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.

Reconstruction Algorithms

Transform-based methods

Formulate the forward problem as a certain *transform*, then formulate a stable way to *invert* the transform.

2D CT: Radon transform $\leftrightarrow$ filtered back projection (FBP).

Works well when we have lots of projection data with low noise.

Algebraic iterative methods

Discretize the forward problem and solve the corresponding large-scale problem $\boxed{A\,x = b}$ by means of an *iterative method*.

A simple example is **Landweber iteration = steepest descent**:

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

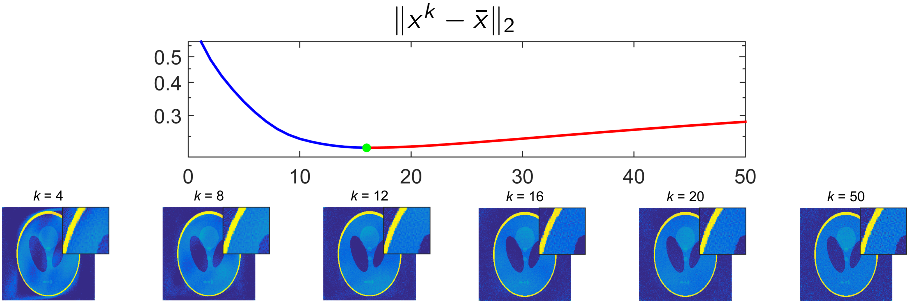
Other examples: SART, Kaczmarz (ART), Landweber-Kaczmarz, CGLS.

Suited for applications with few projections and/or high noise.

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.

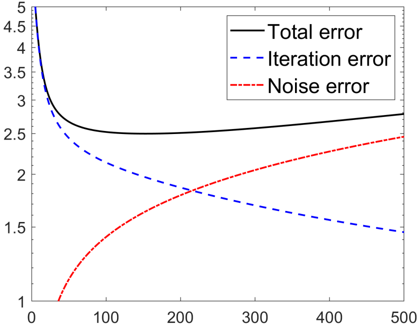
Noise Error and Iteration Error

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

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

Convergence analysis – which we skip here – shows that the iteration error decreases and the noise error increases.

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

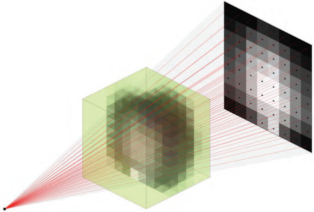


Projectors and Matrices

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 (it is *sparse*), then we use A^T for back projection B, and our stationary iterative methods solve least squares problems associated with the normal equations $A^T A x = A^T b$.

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



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.

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

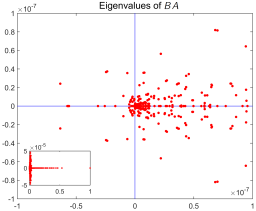
$$0 < \omega < \frac{2\,\Re\mathrm{eal}(\lambda_j)}{|\lambda_j|^2} \quad \text{and} \quad \Re\mathrm{eal}(\lambda_j) > 0, \qquad \{\lambda_j\} = \mathtt{eig}(B\,A)\ .$$

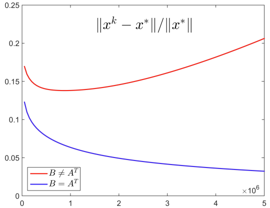
Zeng & Gullberg (2000): similar analysis but ignoring complex λ_j.

Non-Convergence – and Fixing It

Unmatched pair from *ASTRA* software package, image has 64×64 pixels, 90 proj. angles, 60 detector elements, no noise, $\min \Re\text{eal}\, \lambda_j = -6.4 \cdot 10^{-8}$.

No convergence when $B \neq A^T$





Define the **Shifted BA Iteration** (T. Elfving)

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

Convergence condition:

$$\Re\text{eal}\big(\texttt{eig}(B\,A)\big) + \alpha > 0\ .$$

Just choose α large enough that this is satisfied.

Drawback: in addition to relax. param. ω we must also choose shift α.

Alternative: Solve the Unmatched Normal Equations

Instead of "fixing" an iterative method designed for another problem, just solve the **unmatched normal equations** in one of the forms

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

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

$\boxed{M\,x = d}$ 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

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

with the *Krylov subspace*

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

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,
- no need for relaxation parameter or shift parameter,
- fairly simple to implement $\rightarrow$ next page.

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 $i = 1, 2, \ldots, k$
    $h_{i,k} = q_k^T w_i$
    $q_k = q_k - h_{i,k}\,w_i$
  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 $i = 1, 2, \ldots, k$
    $h_{i,k} = q_k^T w_i$
    $q_k = q_k - h_{i,k}\,w_i$
  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

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

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

Properties of the ABBA Algorithms as LS Solvers

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 $\text{range}(B) = \text{range}(A^T)$.

▷ 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 $\text{range}(B^T) = \text{range}(A)$.

▷ equivalent to LSMR when $B = A^T$.

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

. . . but we shall demonstrate success ↬

Theory for Semi-Convergence of GMRES (Technical!)

Some insight has been obtained.

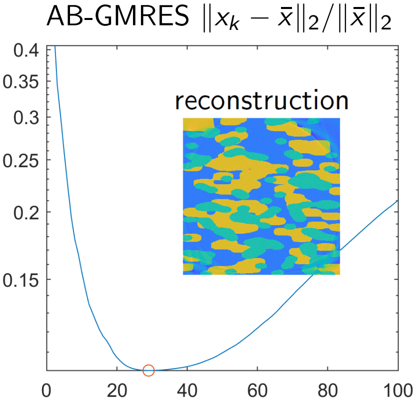
- Calvetti, Lewis, Reichel (2002): if the noise-free data lies in a finite-dimensional Krylov subspace, and if GMRES is equipped with a suitable stopping rule, then the GMRES-solution converges to the exact solution as the noise goes to zero.
- Gazzola, Novati (2016): if the discrete Picard condition (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.

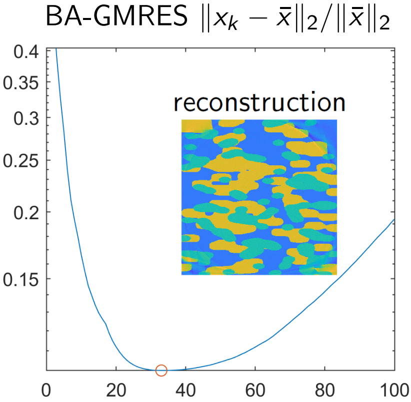
If the SVD components for the large singular values are captured in order of decreasing magnitude, then GMRES will exhibit semi-convergence.

A complete understanding of these aspects has not emerged yet.
Here we rely on insight from numerical experiments (cf. appendix).

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

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





- Semi-convergence is evident for both methods.
- 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:

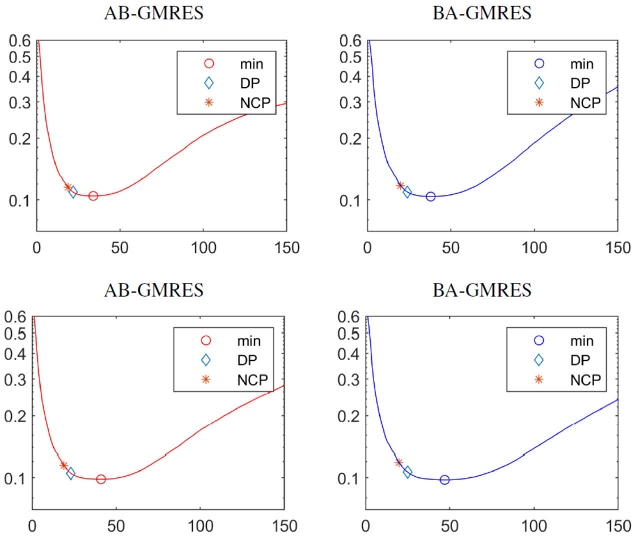
$$k_{\mathrm{DP}} = \text{the smallest } k \text{ for which } \|b - A\,x_k\|_2 \leq \tau\,\|e\|_2$$

  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.

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.

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

New computational experiments by:

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

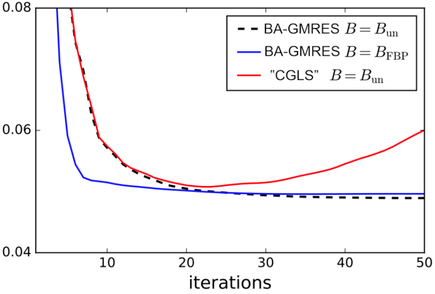
- Cone-beam CT data from an Epica Pegaso veterinary CT scanner.
- 180 projections taken uniformly over one circular rotation.
- Physical "quality assurance" (QA) phantom.
- Detector: 1088×896 pixels of size $(0.278\text{mm})^2$.
- 3D reconstruction: $1024 \times 1024 \times 300$ voxel grid.

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.

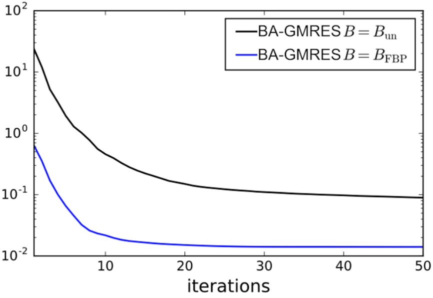
The LS Residual and the UNE Residual

$\|A\,x_k - b\|_2$



Decay: no guarantee

$\|B\,A\,x_k - B\,b\|_2$



Decay: guaranteed

BA-GMRES works well with both B-matrices.

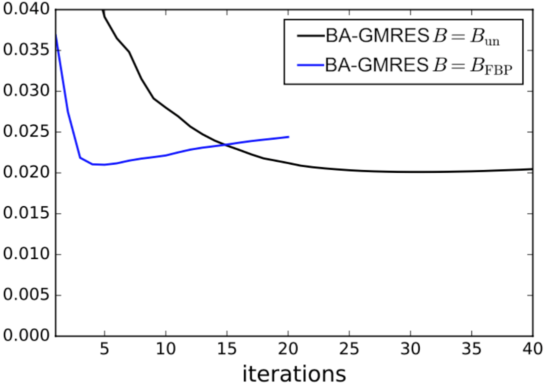
"CGLS" = CGLS with A^T replaced by B; it cannot converge.

"Reconstruction Error"

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

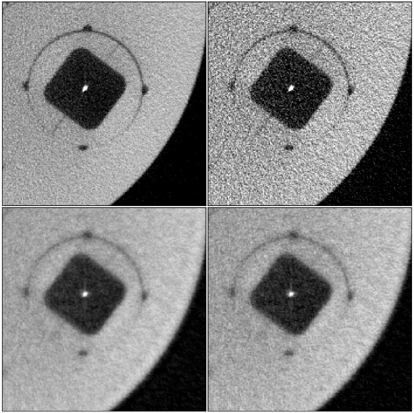
Instead we use a high-quality FBP reconstruction x_{FBP}.

$$\|x_k - x_{\mathrm{FBP}}\|_2$$



With both B-matrices we observe semi-convergence.

Reconstruction, Mid-Slice ROI



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

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.

★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: intuitive understanding of semi-convergence for noisy data.

Future

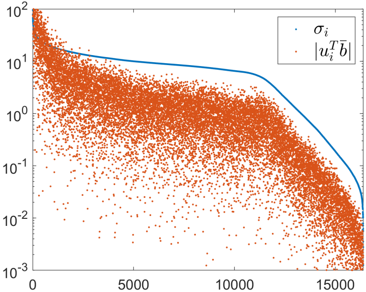
- More theory about semi-convergence for GMRES.
- Deal with memory issue: restart, recycling, etc.
- Ready-to-use implementations for the CT community.

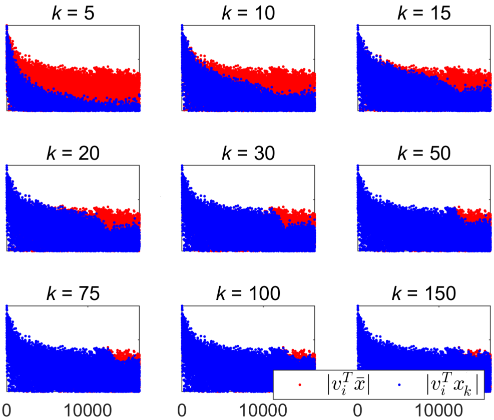
BA-GMRES: SVD Analysis, Small Matrix $23\,040 \times 16\,384$

Right-hand side: $\bar{b} = A\,\bar{x}$

$\bar{x}$ = ground truth

no noise





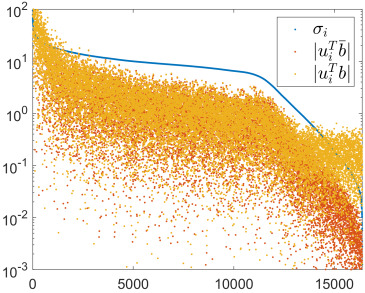
- Left plot is typical for X-ray CT problems; no rank deficiency.
- As k increases we capture more SVD components in x_k.
- At $k = 30$ we already capture the first $11\,000$ exact SVD components.
-
-

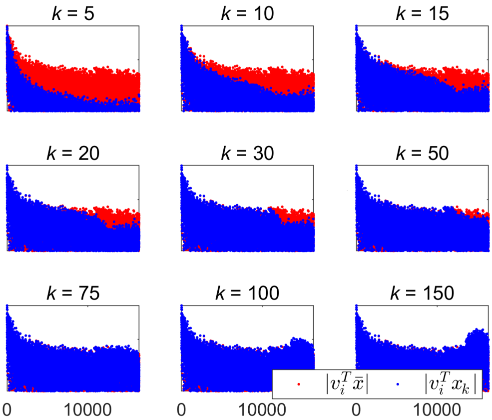
BA-GMRES: SVD Analysis – Now With Noisy Data

Right-hand side: $b = A\,\bar{x} + e$

$\bar{x}$ = ground truth

$\|e\|_2 / \|\bar{b}\|_2 = 0.003$ Gaussian





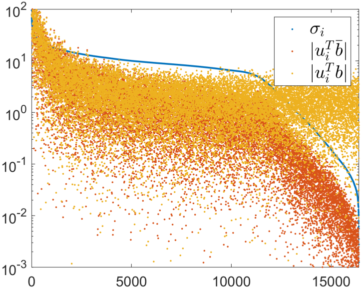
- Left plot is typical for X-ray CT problems; no rank deficiency.
- As k increases we capture more SVD components in x_k.
- At $k = 30$ we already capture the first 11 000 exact SVD components.
- Eventually we include noisy SVD components = **semi-convergence**.
- We obtain the best reconstruction after $k \approx 50$ iterations.

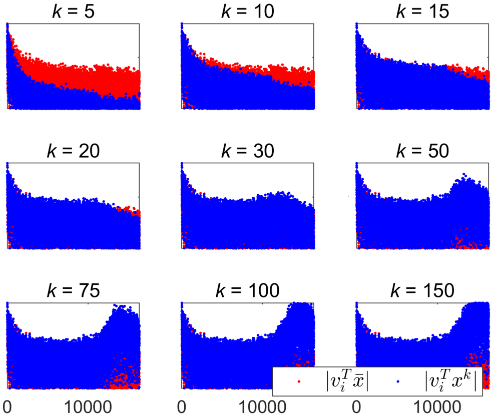
BA-GMRES: SVD Analysis – With More Noise

Right-hand side: $b = A\,\bar{x} + e$

$\bar{x}$ = ground truth

$\|e\|_2 / \|\bar{b}\|_2 = 0.03$ Gaussian





- Left plot: the "noise floor" increases..
- As k increases we capture more SVD components in x_k.
- At $k = 30$ we already capture the first 11 000 exact SVD components.
- Eventually we include noisy SVD components = **semi-convergence**.
- Now we obtain the best reconstruction after $k \approx 20$ iterations.