ART Performance

Why is the Algebraic Reconstruction Technique So Fast?

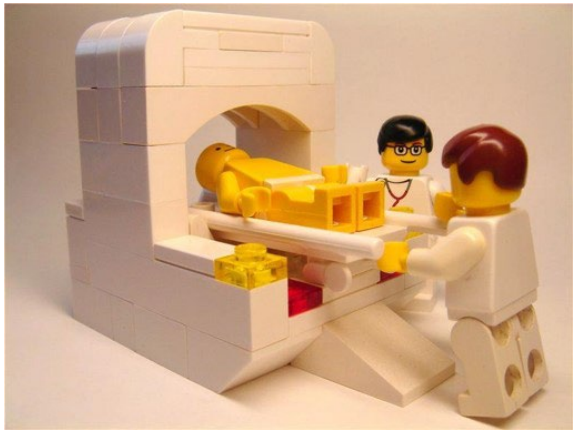
Per Christian Hansen

Joint work with Michiel E. Hochstenbach, TU Eindhoven

DTU Compute
Department of Applied Mathematics and Computer Science

ART = Algbraic Reconstruction Technique

Our motivation: solve linear systems of equations $\boxed{A\,x = b}$ derived from discretization of an underlying tomography problem, e.g., in medical imaging.



ART is a simple iterative method for solving $A\,x = b$ where each iteration updates x via sweeps over the rows a_i^T of the matrix $A \in \mathbb{R}^{m \times n}$.

See the Appendix for ART history.

Among the many iterative solvers, ART has very fast initial convergence $\rightarrow$ making it favorable when only few iterations can be afforded.

ART = Kaczmarz's Method for Solving $Ax = b$

Kaczmarz (1937): repeatedly, orthogonally project x on the hyperplane defined by a_i^T, the ith row of A, and the corresponding element b_i of the right-hand side.

This means that for $k = 0, 1, \ldots$ we carry out:

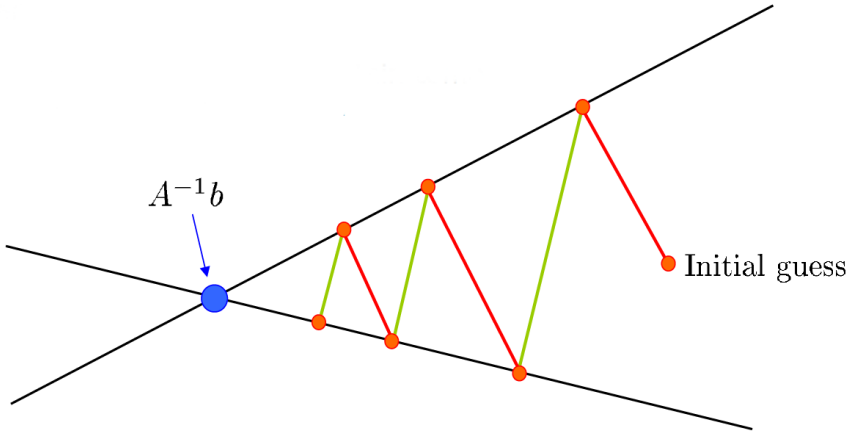
$$\mathbf{x}_{k+1}^{(0)} = \mathbf{x}_k,$$

$$\mathbf{x}_{k+1}^{(i+1)} = \mathbf{x}_{k+1}^{(i)} + \omega \, \frac{\mathbf{b} - \mathbf{a}_i^\top \mathbf{x}_{k+1}^{(i)}}{\|\mathbf{a}_i\|_2^2} \, \mathbf{a}_i, \qquad i = 1, \ldots, m,$$

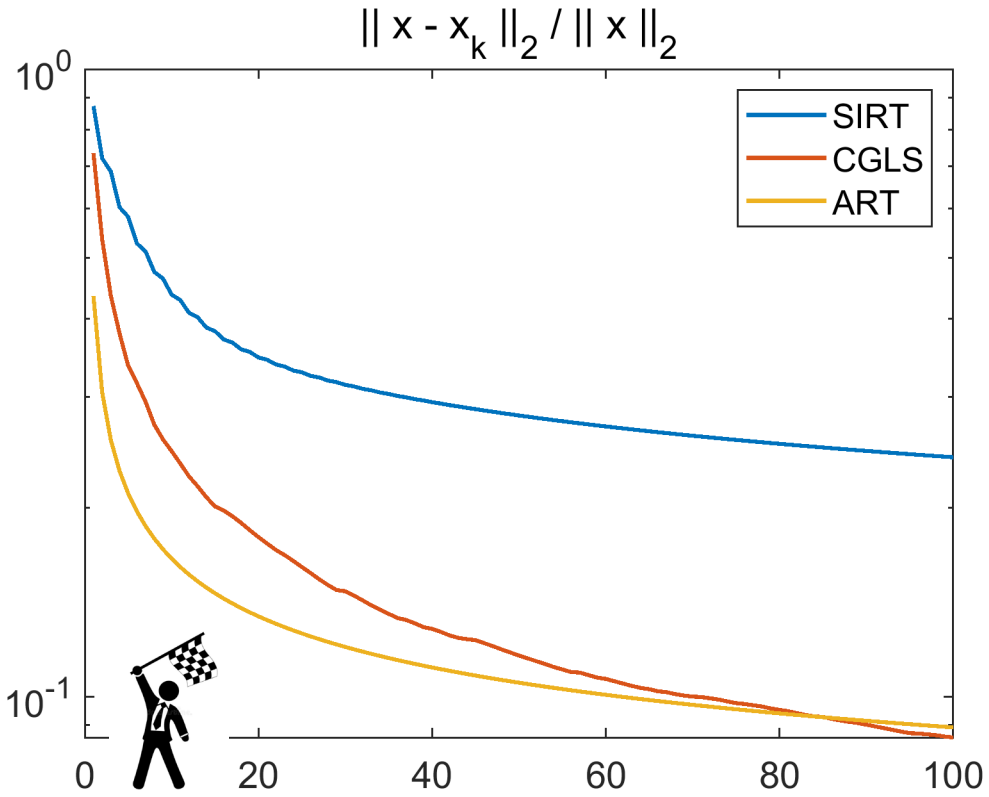
$$\mathbf{x}_{k+1} = \mathbf{x}_{k+1}^{(m)}$$

where ω is a relaxation parameter with $0 < \omega < 2$.

Our starting vector $\mathbf{x}_0 = \mathbf{0}$.



ART Performance



Test problem: **tomo** from Regularization Tools with 100×100 image.

ART is the winner after 10 iterations.

A Bit of Linear Algebra

Decompose $AA^\top = \widehat{L} + D + \widehat{L}^\top$, where $\widehat{L}$ is the strictly lower triangular part and D is the diagonal part of $A\,A^\top$. One sweep of Kaczmarz's method can then be written as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + A^\top L^{-1} \left(\mathbf{b} - A\,\mathbf{x}_k \right) , \qquad L = \widehat{L} + \omega^{-1}\, D \ .$$

Here L is nonsingular (we assume that A has no zero rows).

We can also write the iterative process as

$$\mathbf{x}_{k+1} = G\ \mathbf{x}_k + A^\top\, L^{-1}\, \mathbf{b}$$

with *iteration matrix*

$$G = I - A^\top L^{-1} A \ .$$

Note that G depends on ω through L.

Towards Convergence Studies

Kaczmarz's method is a fixed-point method, and it converges for consistent linear systems and $0 < \omega < 2$. This is equivalent to $\rho(G) < 1$.

We denote the fixed point by $\mathbf{x}_\infty$, and it is given by

$$\mathbf{x}_\infty = (A^\top L^{-1} A)^{-1} A^\top L^{-1} \mathbf{b} \; .$$

Although the iterations – and the speed of convergence – depend on ω, the fixed point $\mathbf{x}_\infty$ is independent of ω. Elfving and Nikazad remind us:

> "there is no underlying function which is minimized."

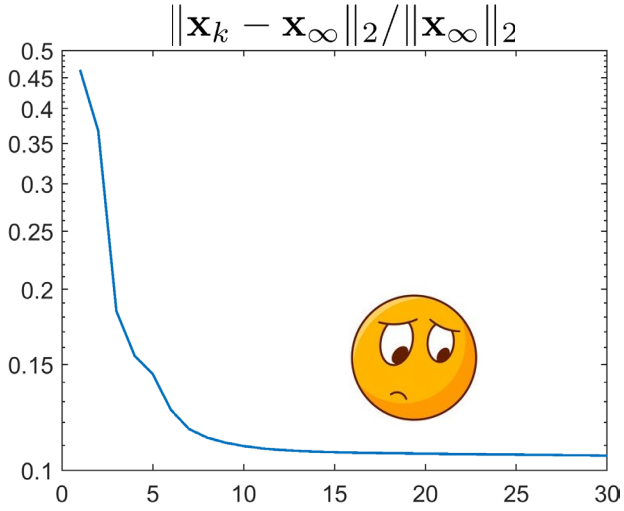
We can rewrite the iteration in the form

$$\mathbf{x}_{k+1} - \mathbf{x}_\infty = (I - A^\top L^{-1} A) \, (\mathbf{x}_k - \mathbf{x}_\infty) = G \, (\mathbf{x}_k - \mathbf{x}_\infty) \; .$$

In terms of the *error vector* $\mathbf{f}_k = \mathbf{x}_k - \mathbf{x}_\infty$ this means

$$\mathbf{f}_{k+1} = (I - A^\top L^{-1} A) \; \mathbf{f}_k = G \, \mathbf{f}_k \; .$$

Asymptotic Convergence



The asymptotic convergence of Kaczmarz's method is very slow!

If $A \in \mathbb{R}^{n \times n}$ then we have (Brezinski, Redivo–Zaglia 2013)

$$\|\mathbf{x}_k - \mathbf{x}_\infty\|_2^2 \leq \left(1 - \frac{1}{n\,\mathrm{cond}(A\,A^\top)}\right) \|\mathbf{x}_{k-1} - \mathbf{x}_\infty\|_2^2$$

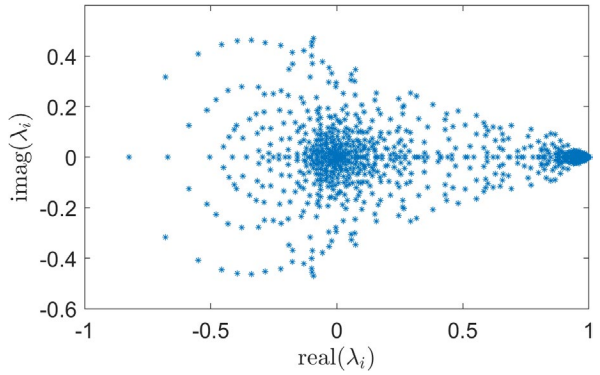
This result was derived earlier for randomized Kaczmarz, where the rows are chosen randomly (Strohmer, Vershynin 2009).

The Role of the Eigenvalues

Recall that the error vector $\mathbf{f}_k = \mathbf{x}_k - \mathbf{x}_\infty$ satisfies $\boxed{\mathbf{f}_{k+1} = G\,\mathbf{f}_k}$.

Hence, we need to scrutinize the eigenvalues of the iteration matrix

$$G = W\,\Lambda\,W^{-1}\,, \qquad \Lambda = \mathrm{diag}(\lambda_i)\,.$$

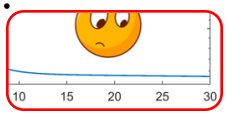


Typical distribution of eigenvalues of *G* for an X-ray CT problem.

Often $\rho(G)$ is extremely close to 1 (and = 1 in finite precision).

Here, $\rho(G)$ = 0.9999998937.

For λ_i close to 1, the corresponding modes (eigenvector components) of $\mathbf{f}_k$ are damped very little in each iterations.

This explains the asymptotic "plateau" 

for the iteration error.

For $\lambda_i \approx 0$, the corresponding eigenvector components of $\mathbf{f}_k$ vanish fast.

A Zero Eigenvalue of *G*

We now consider the important case of $\omega = 1$, which is often default.

Then $G = I - A^\top L^{-1} A$ always has a zero eigenvalue with corresponding eigenvector $\mathbf{a}_1 = A^\top \mathbf{e}_1$ (the transpose of the first row of A).

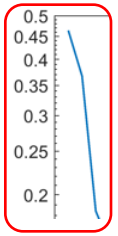
Proof. Recall $A\,A^\top = L + \widehat{L}^\top$ with L lower triangular and $\widehat{L}^\top$ strictly upper triangular. Hence,

$$A\,\mathbf{a}_1 = A\,A^\top \mathbf{e}_1$$

and

$$G\,\mathbf{a}_1 = (I - A^\top L^{-1} A)\,A^\top \mathbf{e}_1 = A^\top \mathbf{e}_1 - A^\top \mathbf{e}_1 = 0\ .$$

A zero eigenvalue of G means that the corresponding mode converges after one iteration. Hence, if $\mathbf{x}_\infty$ has a large component in the direction of $\mathbf{a}_1$ then Kaczmarz makes rapid immediate progress.



More Zero Eigenvalues of *G*

We still consider $\omega = 1$, and study when G may have more zero eigenvalues.

This happens when A has *structurally orthogonal* rows with $\mathbf{a}_i^\top \mathbf{a}_j = 0$.

To see this, let $\mathbf{e}_j$ denote the jth unit vector, and recall that L is the lower traingular part of $A\,A^\top$. Then

$$A\,\mathbf{a}_j = A\,A^\top \mathbf{e}_j = L\,\mathbf{e}_j + \sum_{i<j} \left(\mathbf{a}_i^\top \mathbf{a}_j\right)\,\mathbf{e}_j \ .$$

Specifically, for $j = 2$ we have

$$A\,\mathbf{a}_2 = A\,A^\top \mathbf{e}_2 = L\,\mathbf{e}_2 + \left(\mathbf{a}_1^\top \mathbf{a}_2\right)\,\mathbf{e}_2 \ .$$

When $\mathbf{a}_1^\top \mathbf{a}_2 = 0$ we have $A\,\mathbf{a}_2 = L\,\mathbf{e}_2$ and thus

$$G\,\mathbf{a}_2 = A^\top \mathbf{e}_2 - A^\top L^{-1} A\,A^\top \mathbf{e}_2 = A^\top \mathbf{e}_2 - A^\top \mathbf{e}_2 = 0 \ .$$

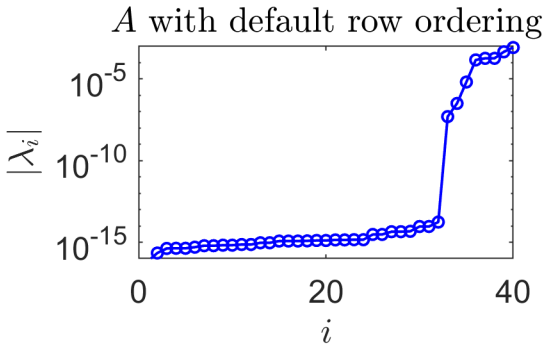
Similarly, $\mathbf{a}_3$ is an eigenvector with $\lambda_3 = 0$ when $\mathbf{a}_1^\top \mathbf{a}_3 = \mathbf{a}_2^\top \mathbf{a}_3 = 0$.

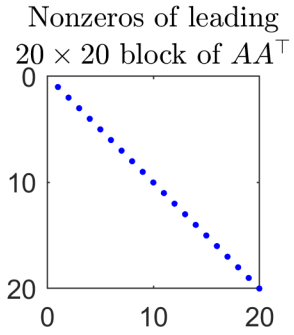
An so forth ...

Example from X-Ray CT

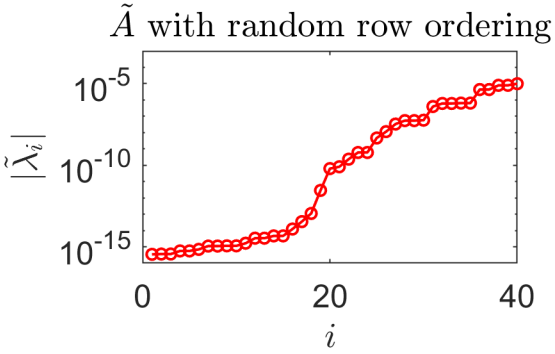
Test problem **paralleltomo** from AIR TOOLS II: 32×32 image, 32 projection angles, 32 X-rays per angle; A is 1024×1024 and has full rank.

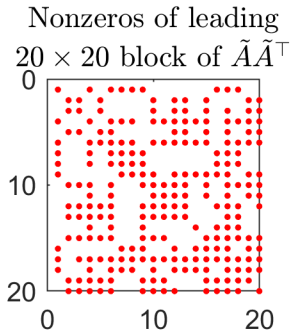
The corresponding matrix G has several (numerically) zero eigenvalues.





The 32 top rows of A are mutually structurally orthogonal, and they are null vectors of G.





Only rows 1 and 2 are structurally orthogonal; these are the only rows of A that are null vectors of G.

Near-Zero Eigenvalues of *G*

When does G may have near-zero eigenvalues?

And when $|\mathbf{a}_1^\top \mathbf{a}_2|$ is nonzero but small, then $\mathbf{a}_2$ is an *approximate* eigenvector of G corresponding to an eigenvalue close to 0. Here is why:

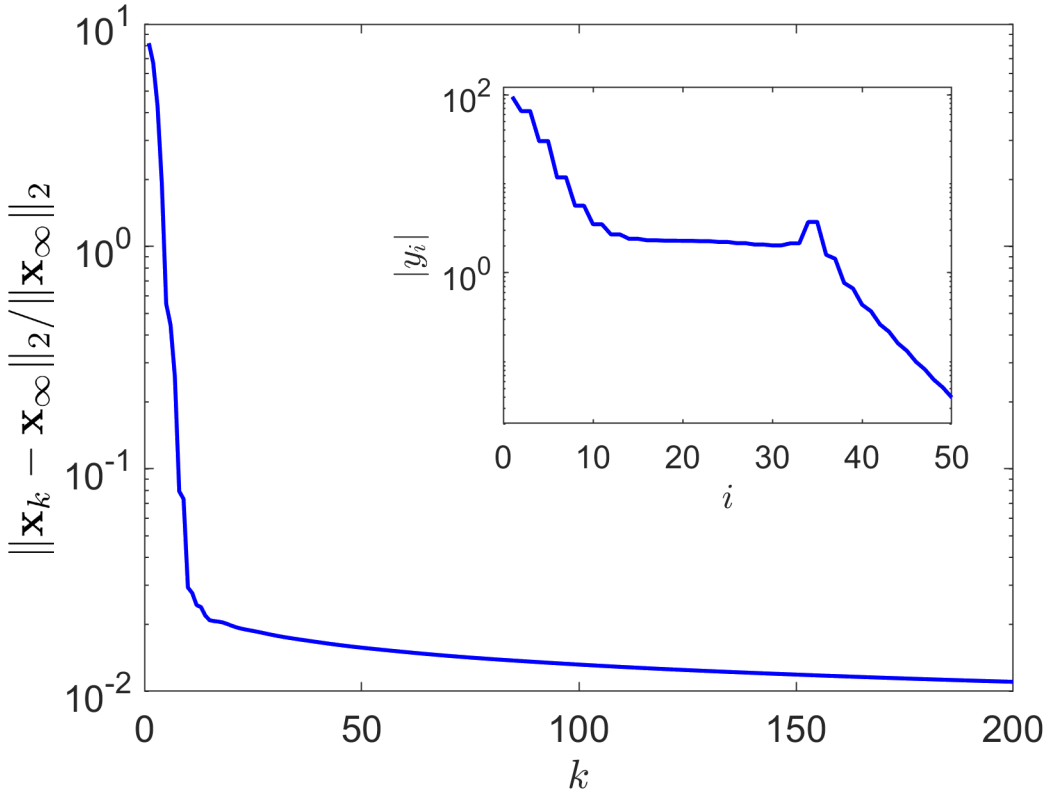
If $L^{-1} A\, A^\top$ is diagonalizable, the Bauer–Fike Theorem guarantees the existence of an eigenvalue λ of G with

$$|\lambda| \leq \kappa(X) \cdot |\mathbf{a}_1^\top \mathbf{a}_2| \cdot \|L^{-1}\mathbf{e}_1\|_2 \ ,$$

where $\kappa(X)$ is the condition number of the eigenvector matrix X of $L^{-1} A\, A^\top$.

Good row orderings may therefore have favorable effects, not only for the asymptotic convergence (smaller $\rho(G)$), but also for the initial stage (near-zero eigenvalues).

Another Example: `gravity` from Reg. Tools



n = 128

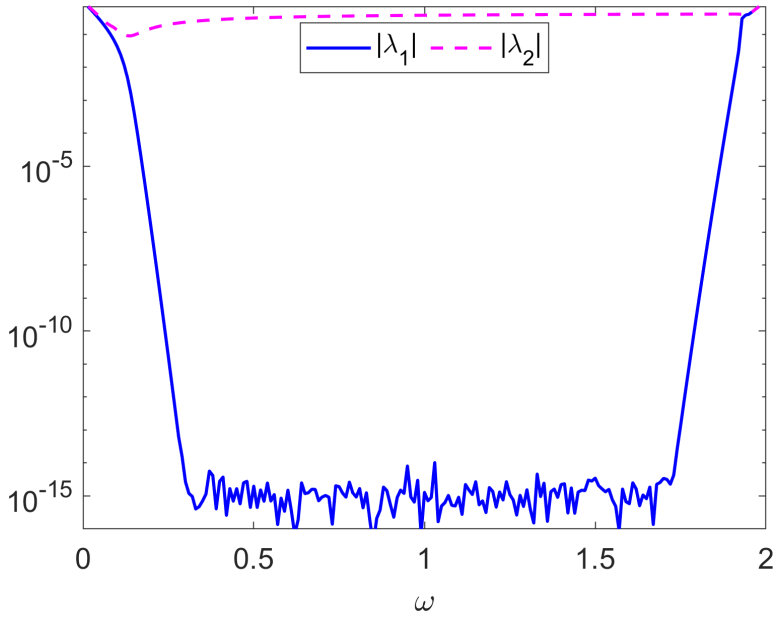
The eigenvalues of G are sorted such that $0 = \lambda_1 \leq |\lambda_2| \leq |\lambda_3| \leq \cdots$ and y_i are the elements of $\mathbf{y} = W^{-1}\mathbf{x}_\infty$ (coefficients in the eigenvector basis). We have fast initial convergence because $\mathbf{x}_\infty$ is dominated by the eigenvectors corresponding to the smallest eigenvalues.

(Near) Zero Eigenvalues when $\omega \neq 1$

When $\omega \neq 1$ it is not guaranteed that there is a zero eigenvalue of G.

But the examples below show that we can have one or more zero eigenvalues for a wide range of ω around 1.

gravity



n = 128

paralleltomo

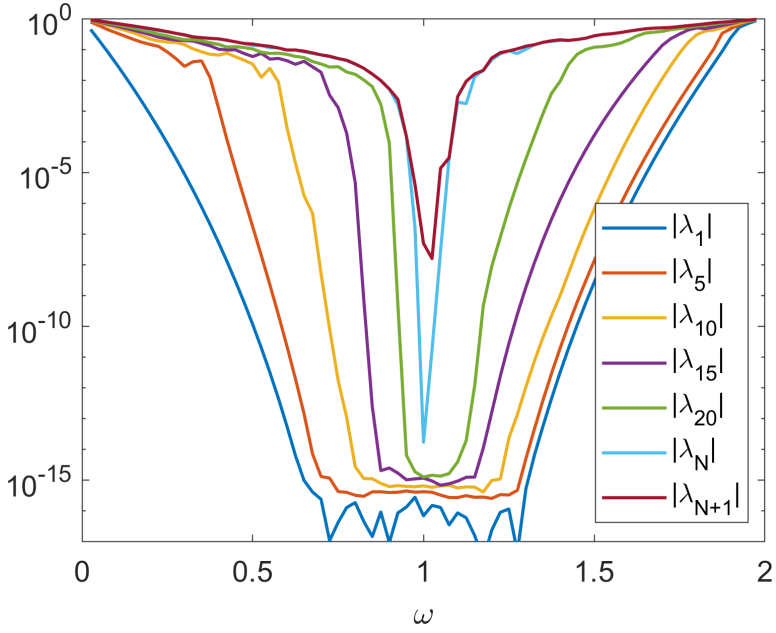
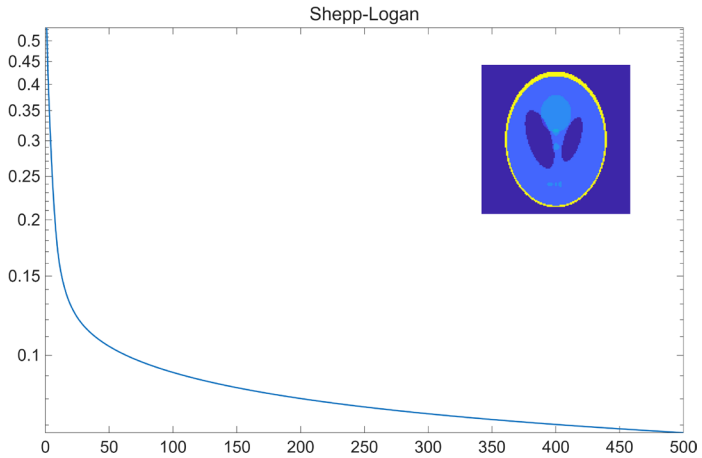
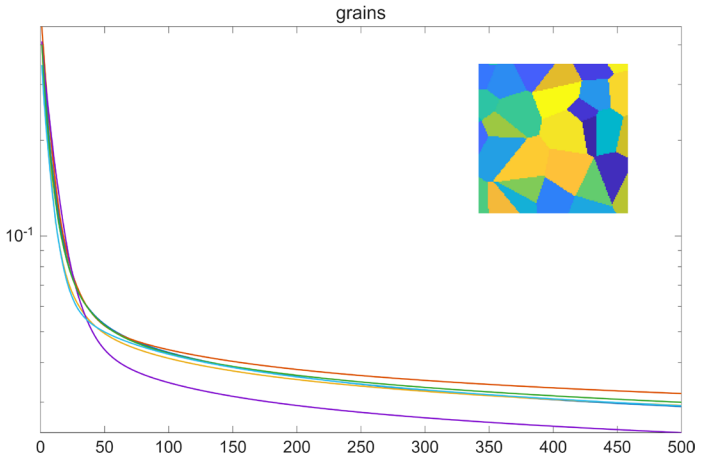


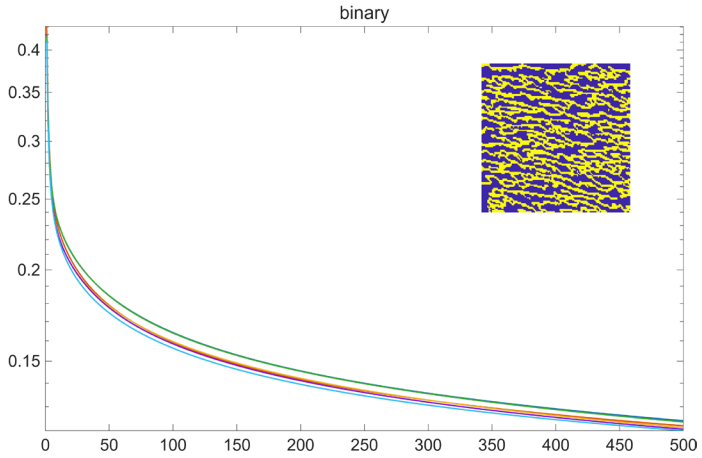
Image is $N \times N$; N = 32; n = 1024

More Experiments with `paralleltomo`

Convergence histories for four different test images (phantoms).
The last three are random, and we show results for five different seeds.







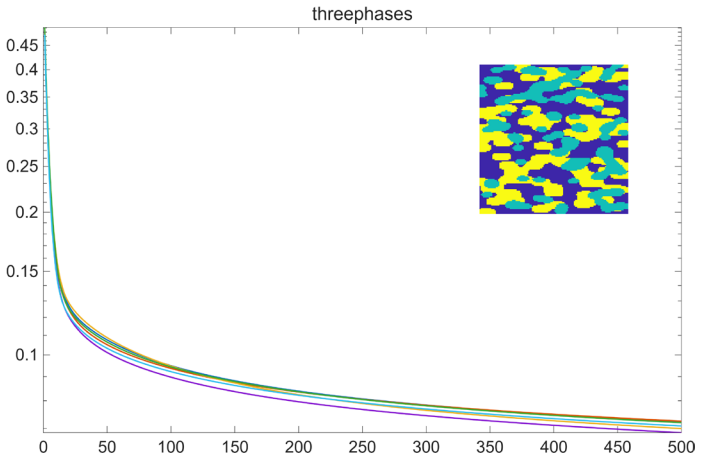
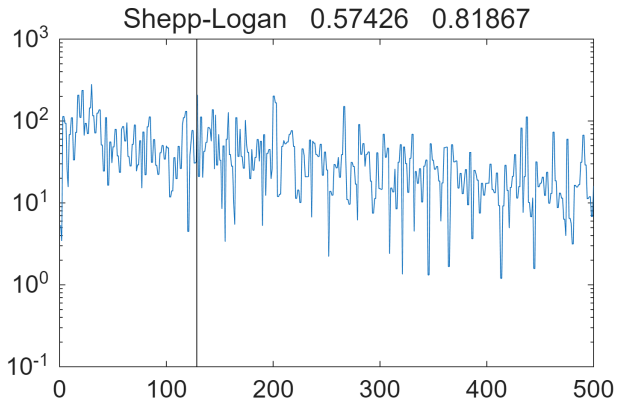


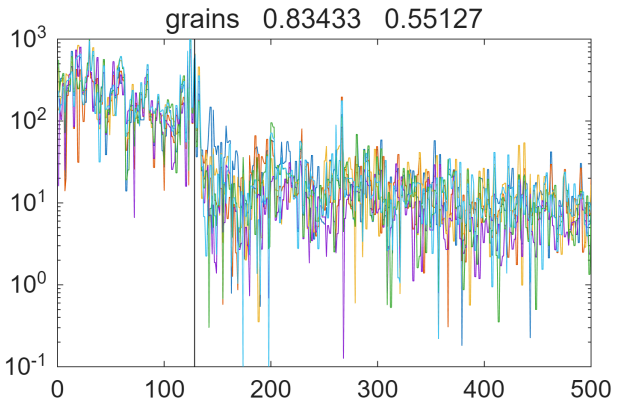
Image is $N \times N$.
N projection angles.
N detectors (rays).
$N = 128$.

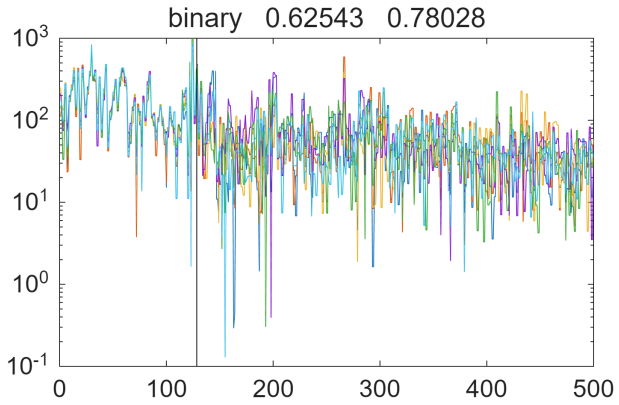
In all cases, we see a fast initial convergence – as expected.

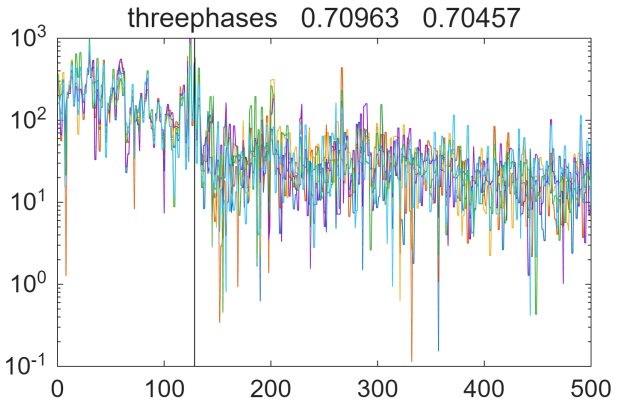
Eigenvector Components

Abs. values of first 500 elements of $\mathbf{y} = W^{-1}\mathbf{x}_\infty$ (the coefficients in the eigenvector basis). The two numbers are $\|P_0\,\mathbf{x}_\infty\|_2$ and $\|(I - P_0)\,\mathbf{x}_\infty\|_2$, where P_0 = orthogonal projector unto the 128-dim. null space of G.









In all cases, a substantial component of $\mathbf{x}_\infty$ lies in the null space.

BONUS – Noise Propagation: Set the Stage

We consider white Gaussian noise in the right-hand side:

$$\mathbf{b} = \bar{\mathbf{b}} + \mathbf{e} \;, \qquad \bar{\mathbf{b}} = A\,\bar{\mathbf{x}} \;, \qquad \mathbf{e} \sim \mathcal{N}(0,\, \sigma^2 I) \;,$$

where $\bar{\mathbf{x}}$ is the exact solution (the ground truth), $\bar{\mathbf{b}}$ is the noise-free data, $\mathbf{e}$ is the noise, and σ is its element-wise standard deviation.

With $A_k^{\#} = (I - G^k)A^{\#}$ we can write the kth iterate as

$$\mathbf{x}_k = A_k^{\#}\,\mathbf{b} = \bar{\mathbf{x}}_k + A_k^{\#}\,\mathbf{e} \qquad \text{where} \qquad \bar{\mathbf{x}}_k = A_k^{\#}\,\bar{\mathbf{b}} \;;$$

we refer to $\bar{\mathbf{x}}_k$ as the noise-free iterates. We then split* the error:

$$\mathbf{x}_k - \bar{\mathbf{x}} = \underbrace{\mathbf{x}_k - \bar{\mathbf{x}}_k}_{\text{noise error}} + \underbrace{\bar{\mathbf{x}}_k - \bar{\mathbf{x}}}_{\text{it. error}} \;.$$

Here, $\bar{\mathbf{x}}_k - \bar{\mathbf{x}}$ is the *iteration error*. The other component $\mathbf{x}_k - \bar{\mathbf{x}}_k = A_k^{\#}\,\mathbf{e}$ is the *noise error*, and it describes the propagated noise from the data.

---

* Such splittings go back to Aulick & Gallie (1983), and probably further back.

Noise Propagation: Example

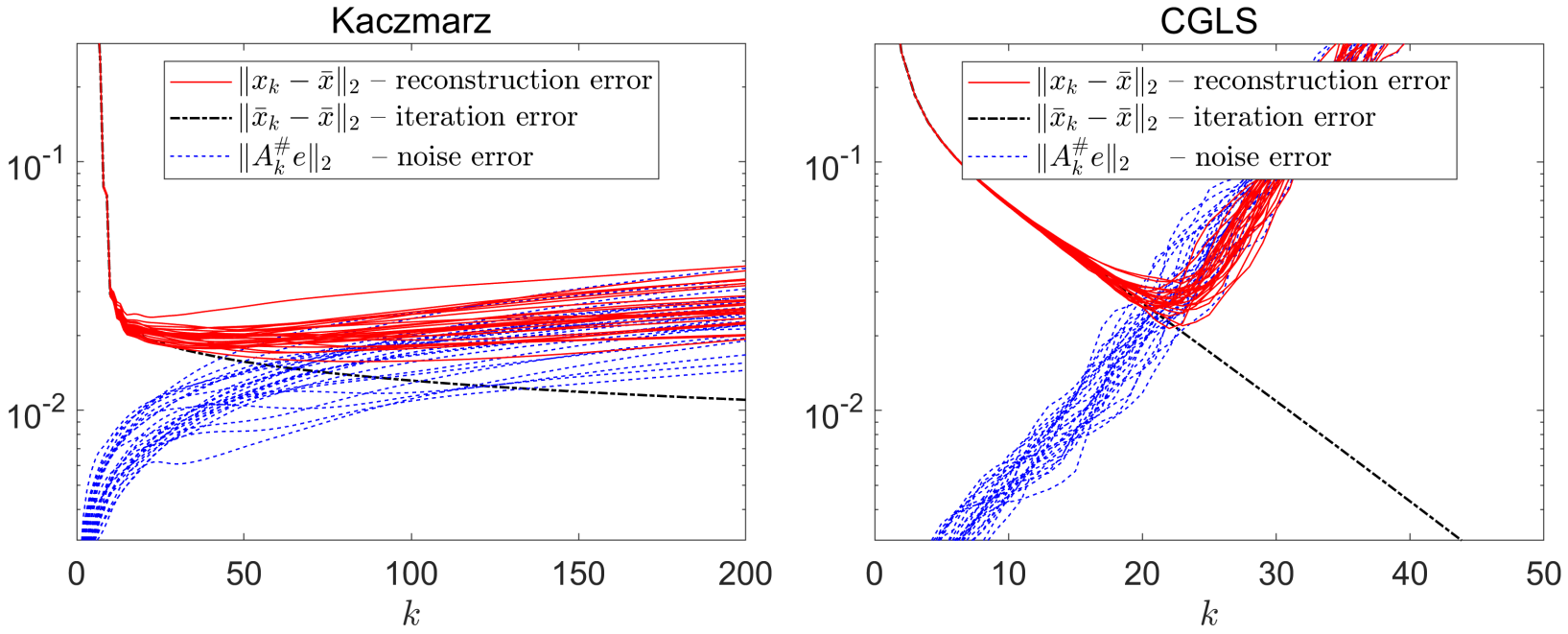


Illustration of the three errors for the gravity test problem, and with 25 realization of the noise.

The results for CGLS – whose behavior is analyzed in (Hansen 2025) – show that the underlying behavior of semi-convergence is the same for both methods.

Noise Propagation: the Noise Error

Our aim is to study the statistics of $\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|_2 = \|A_k^{\#}\,\mathbf{e}\|_2$.

It is convenient to do this in the eigenvector basis:

$$\mathbf{x}_k - \bar{\mathbf{x}}_k = A_k^{\#}\mathbf{e} = (I - G^k)\,A^{\#}\mathbf{e} = W\,(I - \Lambda^k)\,W^{-1}A^{\#}\mathbf{e}\ .$$

If we introduce new variables

$$\boldsymbol{\xi}^k = W^{-1}A_k^{\#}\,\mathbf{e}\ , \qquad \boldsymbol{\xi} = W^{-1}A^{\#}\,\mathbf{e}$$

then we have

$$\boldsymbol{\xi}^k = (I - G^k)\,\boldsymbol{\xi}\ .$$

Then it follows that

$$\left\|\boldsymbol{\xi}^k\right\|_2^2 = \left\|(I - G^k)\,\boldsymbol{\xi}\right\|_2^2 = \sum_{i=1}^{n} \left|1 - \lambda_i^k\right|^2\ |\xi_i|^2\ ,$$

where ξ_i are the elements of $\boldsymbol{\xi}$; k only enters via the factor $I - G^k$.

Cannot guarantee that $\|A_k^{\#}\mathbf{e}\|_2$ or $\|\boldsymbol{\xi}^k\|_2$ increases monotonically with k.

Statistics of the Noise Error

Theorem. The expected value related to the noise error satisfies

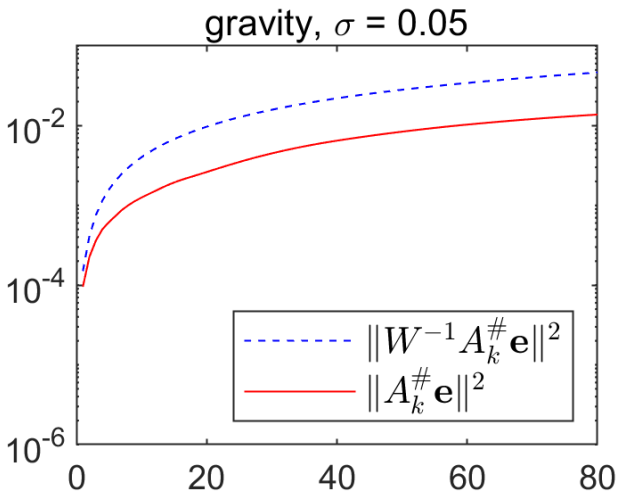
$$\mathbb{E}(\|A_k^{\#}\,\mathbf{e}\|_2^2) = \sigma^2\ \|A_k^{\#}\|_{\mathrm{F}}^2 = \sigma^2\ \|W\,(I-\Lambda^k)\,W^{-1}A^{\#}\|_{\mathrm{F}}^2$$

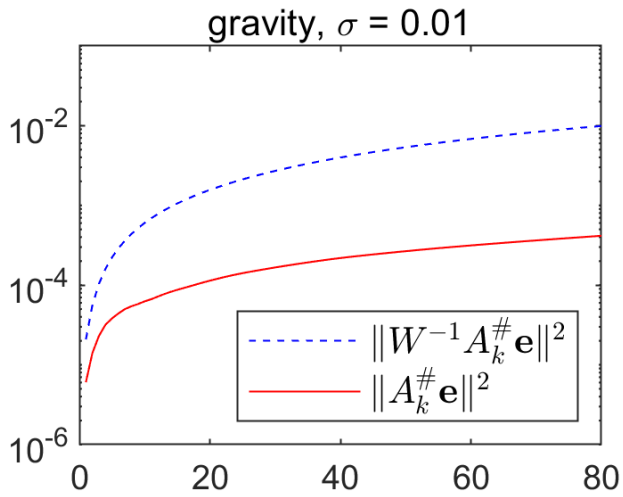
and

$$\mathbb{E}(\|\boldsymbol{\xi}^k\|_2^2) = \sum_{i=1}^{n}\ |1-\lambda_i^k|^2\ \mathbb{E}(|\xi_i|^2)\ .$$

This is where I am stuck, because these expressions are hard to handle.

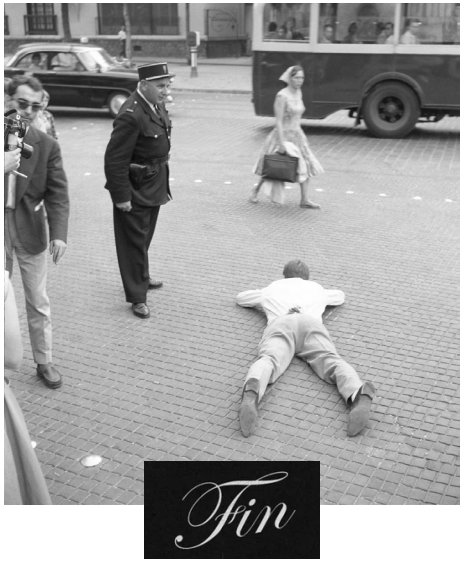
Expected values.





Checking Out

1. We pointed out the need for analysis of the initial convergence.
2. We give new insight about (near) zero eigenvalues of iteration matrix.
3. We explain the fast initial convergence when the solution is rich in the modes correspoding to zero and small eigenvalues.
4. Often true in X-ray CT and other applications.
5. Need more analysis of the error propagation.



P. C. Hansen and M. E. Hochstenbach, *On spectral properties and fast initial convergence of the Kaczmarz method*, BIT Numerical Mathematics (invited paper), 66 (2026), paper 8, doi 10.1007/s10543-025-01098-1.

Appendix: ART History

Kaczmarz (1937): cyclic sweeps:

$$\mathbf{x} \leftarrow \mathbf{x} + \frac{\mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}}{\|\mathbf{a}_i\|_2^2} \, \mathbf{a}_i \ , \qquad i = 1, 2, \ldots, m \ .$$

Gordon, Bender, Herman (1970): coined the term "ART" and introduced a nonnegativity projection:

$$\mathbf{x} \leftarrow \mathcal{P}_{\mathbb{R}^n_+} \left(\mathbf{x} + \frac{\mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}}{\|\mathbf{a}_i\|_2^2} \, \mathbf{a}_i \right) \ , \qquad i = 1, 2, \ldots, m \ .$$

Herman, Lent, Lutz (1978): introduced relaxation parameters $\omega_k < 2$:

$$\mathbf{x} \leftarrow \mathbf{x} + \omega_k \, \frac{\mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}}{\|\mathbf{a}_i\|_2^2} \, \mathbf{a}_i \ , \qquad i = 1, 2, \ldots, m \ .$$

Today ART includes both ω_k and a projection $\mathcal{P}_C$ on a convex set:

$$\mathbf{x} \leftarrow \mathcal{P}_C \left(\mathbf{x} + \omega_k \, \frac{\mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}}{\|\mathbf{a}_i\|_2^2} \, \mathbf{a}_i \right) \ , \qquad i = 1, 2, \ldots, m \ .$$