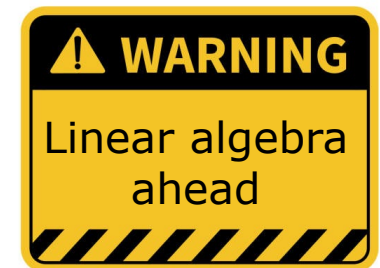


ART Performance

Why is the Algebraic Reconstruction Technique So Fast?

Per Christian Hansen

Joint work with Michiel E. Hochstenbach, TU Eindhoven



ART = A *Classical* Algorithm

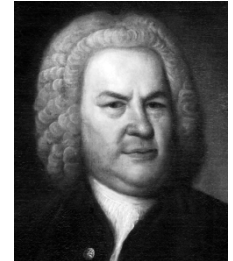
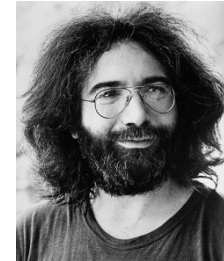
Perspective:

Listen to Grateful Dead (1965–1995) → old fashioned.

Listen to Mozart (1756–91) or Bach (1685–28) → the classics!

Talk about total variation (1992) → old stuff.

Talk about ART (1937) → classical algorithm.



Our motivation: solve linear systems of equations $Ax = b$ derived from discretization of an underlying tomography problem.

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

Among the many iterative solvers, ART has very fast initial convergence → 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 i th row of A , and the corresponding element b_i of the right-hand side.

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

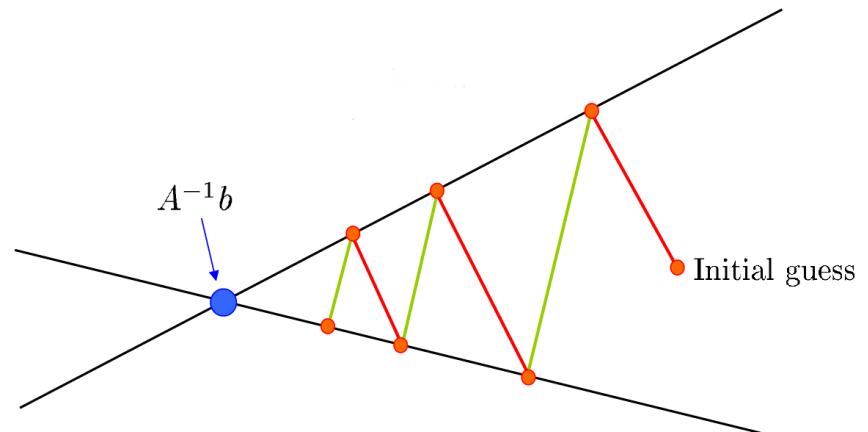
$$\mathbf{x}_{k+1}^{(0)} = \mathbf{x}_k,$$

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

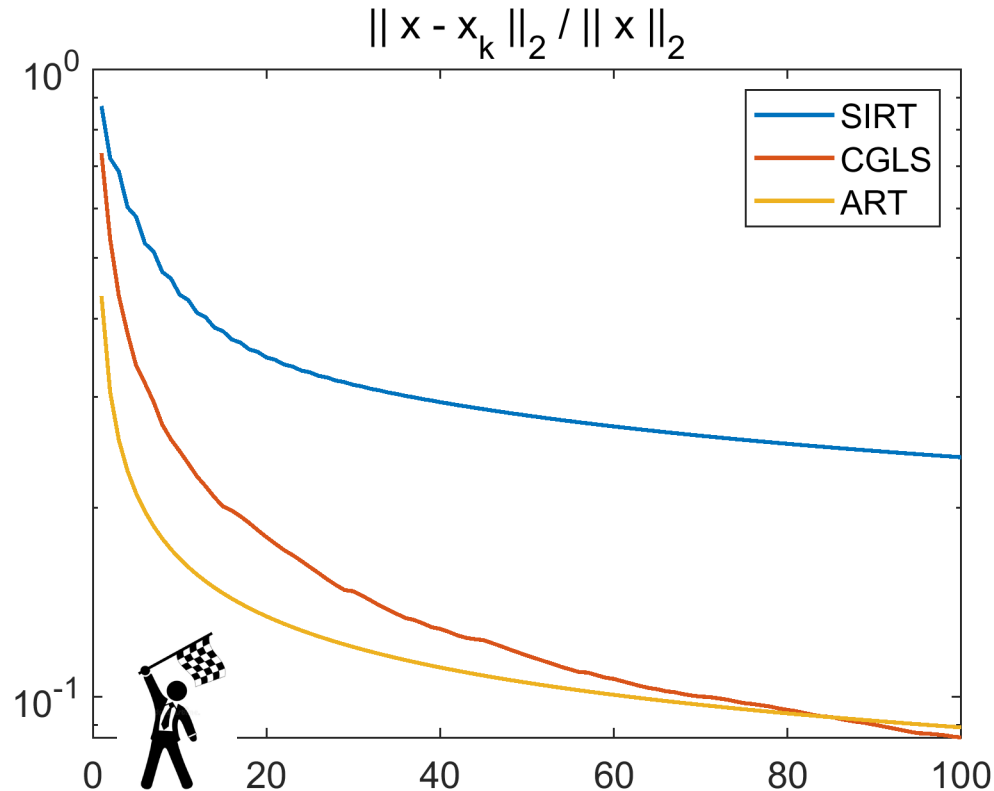
$$\mathbf{x}_{k+1} = \mathbf{x}_{k+1}^{(m)}$$

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

$$\mathbf{x}_{k+1} = \mathbf{x}_k + A^\top L^{-1} (\mathbf{b} - A \mathbf{x}_k), \quad L = L_\omega = \hat{L} + \omega^{-1} D.$$

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. This is equivalent to $\rho(G) < 1$.

We denote the fixed point by \mathbf{x}_∞ , 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}_∞ is independent of ω . Elfving and Nikazad remind us:

“there is no underlying function which is minimized.”

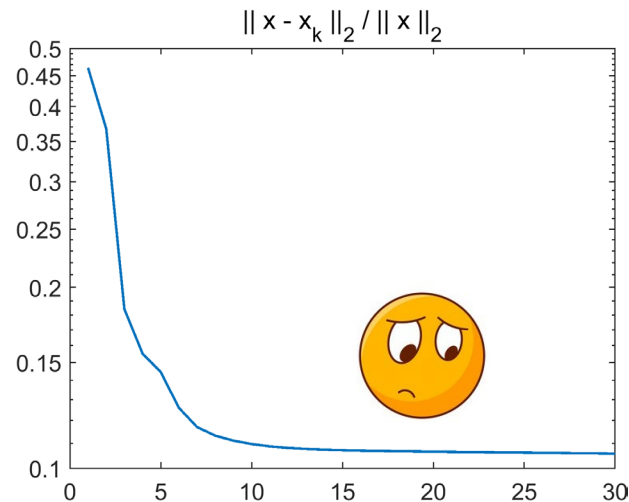
We can rewrite the iteration in the form

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

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

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

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)

$$\|x_k - x_\infty\|_2^2 \leq \left(1 - \frac{1}{n \operatorname{cond}(A A^\top)}\right) \|x_{k-1} - x_\infty\|_2^2$$

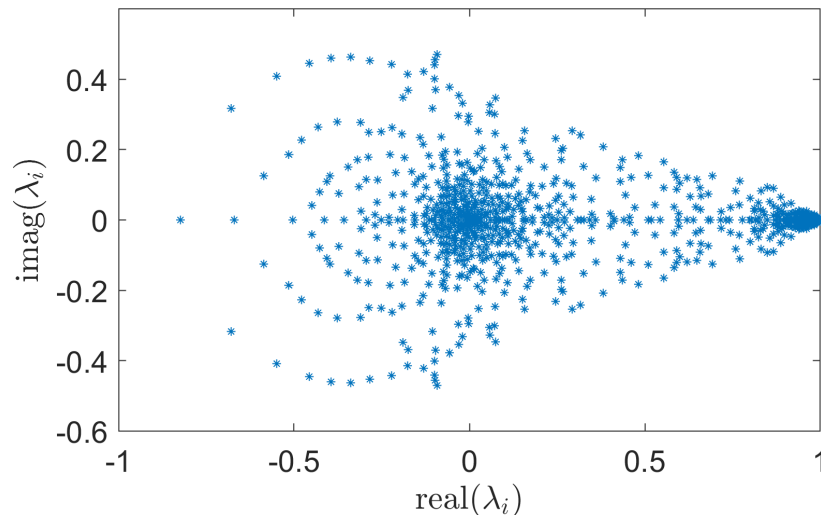
This result was derived earlier for randomized Kaczmarz, where the rows are chosen randomly (Strohmer, Vershynin 2009).

Towards a Study of the Initial Convergence

Clearly, there is a need to study the initial convergence!

To do this we need to scrutinize the eigenvalue of the iteration matrix

$$G = I - A^T L^{-1} A = W \Lambda W^{-1}, \quad \Lambda = \text{diag}(\lambda_i).$$



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

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

We find that $\rho(G)$ may be extremely close to 1 (often equal to 1 when computed in finite precision). This explains the asymptotic plateau for the iteration error.

Zero Eigenvalues of G

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

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

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

Often G has many zero eigenvalues. To see this, let \mathbf{e}_j denote the j unit vector, and recall that L is the lower triangular part of AA^\top . Then

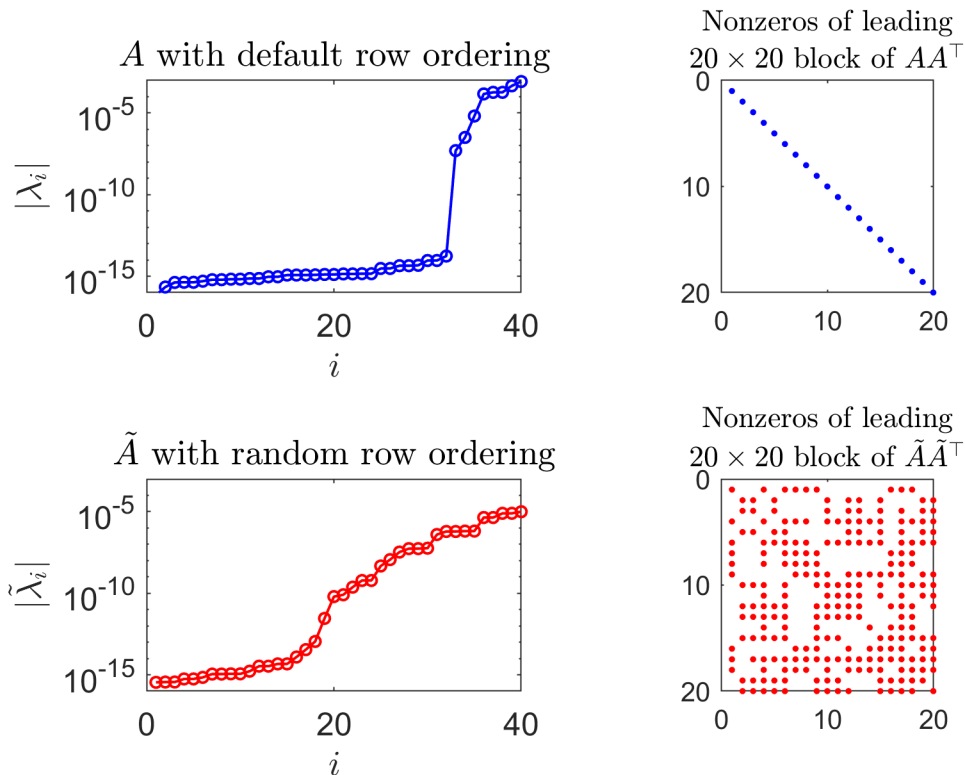
$$AA^\top \mathbf{e}_j = L \mathbf{e}_j + \sum_{i < j} (\mathbf{a}_i^\top \mathbf{a}_j) \mathbf{e}_j .$$

Often, several rows of A are *structurally orthogonal* with $\mathbf{a}_i^\top \mathbf{a}_j = 0$.

- \mathbf{a}_2 is an eigenvector corresponding to eigenvalue zero when $\mathbf{a}_1^\top \mathbf{a}_2 = 0$,
- \mathbf{a}_3 is a such as eigenvector when $\mathbf{a}_1^\top \mathbf{a}_3 = \mathbf{a}_2^\top \mathbf{a}_3 = 0$,
- etc.

Example from X-Ray CT

Test problem `paralleltomo` from AIR TOOLS II: 32×32 image, 32 projection angles, and 32 X-rays per angle; A is 1025×1024 and has full rank. The corresponding matrix G has several (numerically) zero eigenvalues.



Near Zero Eigenvalues of G

When does G may have near-zero eigenvalues?

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:

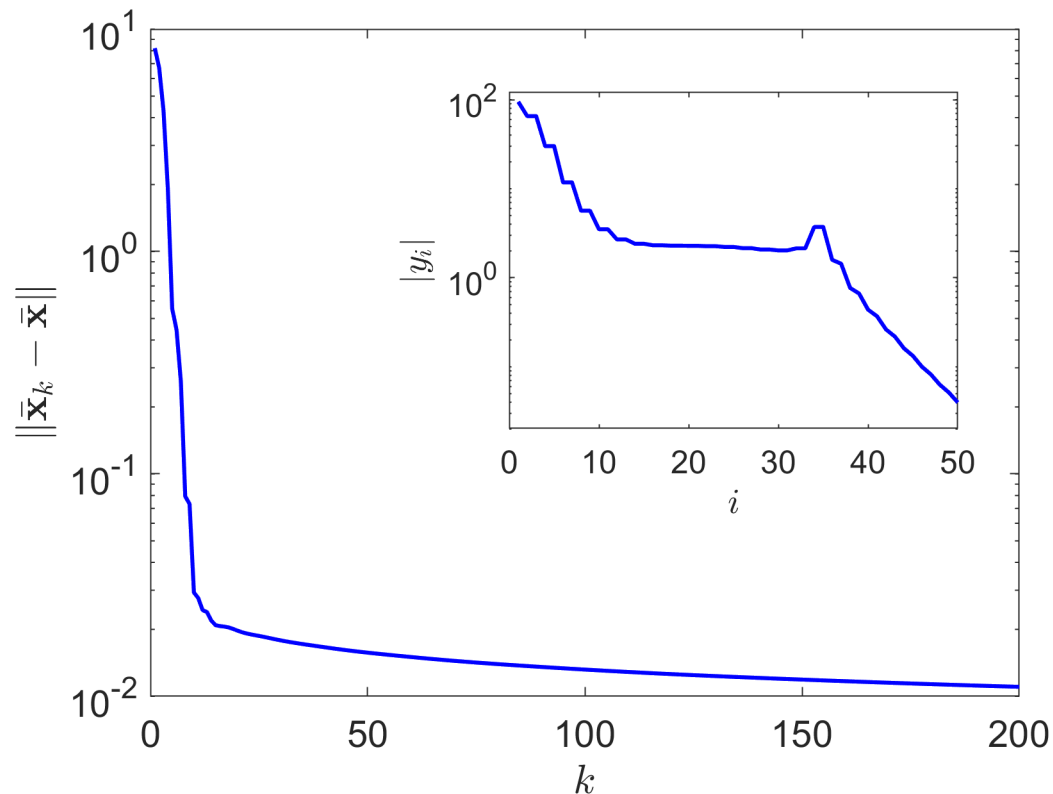
If $L^{-1}AA^\top$ is diagonalizable, the Bauer–Fike Theorem guarantees the existence of an eigenvalue λ of G with

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

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

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

Another Example: gravity from Reg. Tools

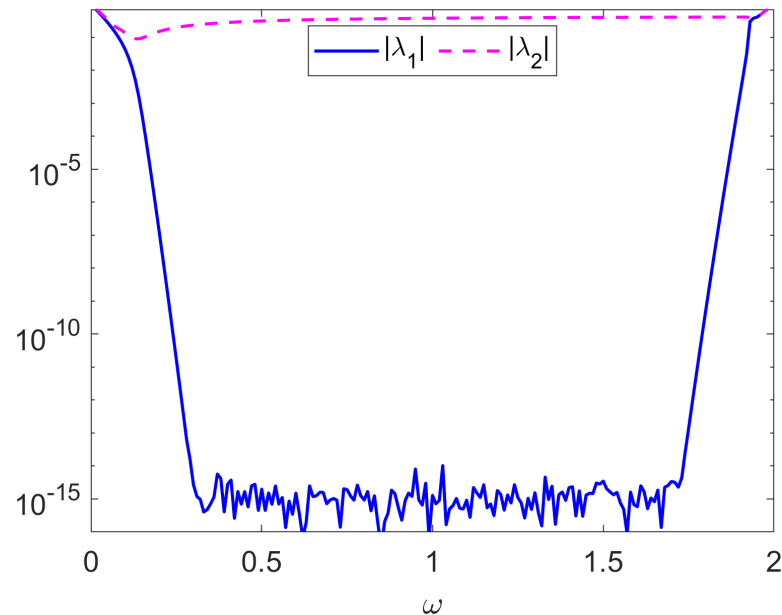


The eigenvalues of G are sorted such that $0 = \lambda_1 \leq |\lambda_2| \leq |\lambda_3| \leq \dots$ and y_i are the elements of $\mathbf{y} = W^{-1}\mathbf{x}_\infty$ (coefficients in the eigenvector basis). We have fast initial convergence because \mathbf{x}_∞ is dominated by the eigenvectors corresponding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

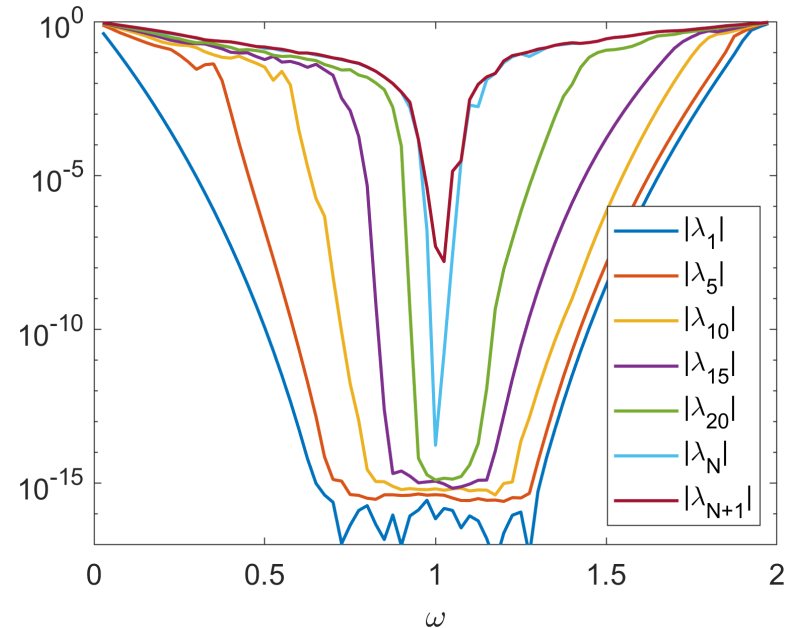
(Near) Zero Eigenvalues when $\omega \neq 1$

When $\omega \neq 1$ it is not guaranteed that there is a zero eigenvalue of G .
 The examples below show that we can have one or more zero eigenvalues for a wide range of ω around 1.

gravity



paralleltomo



BONUS - Noise Propagation: Set the Stage

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

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

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 k th iterate as

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

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

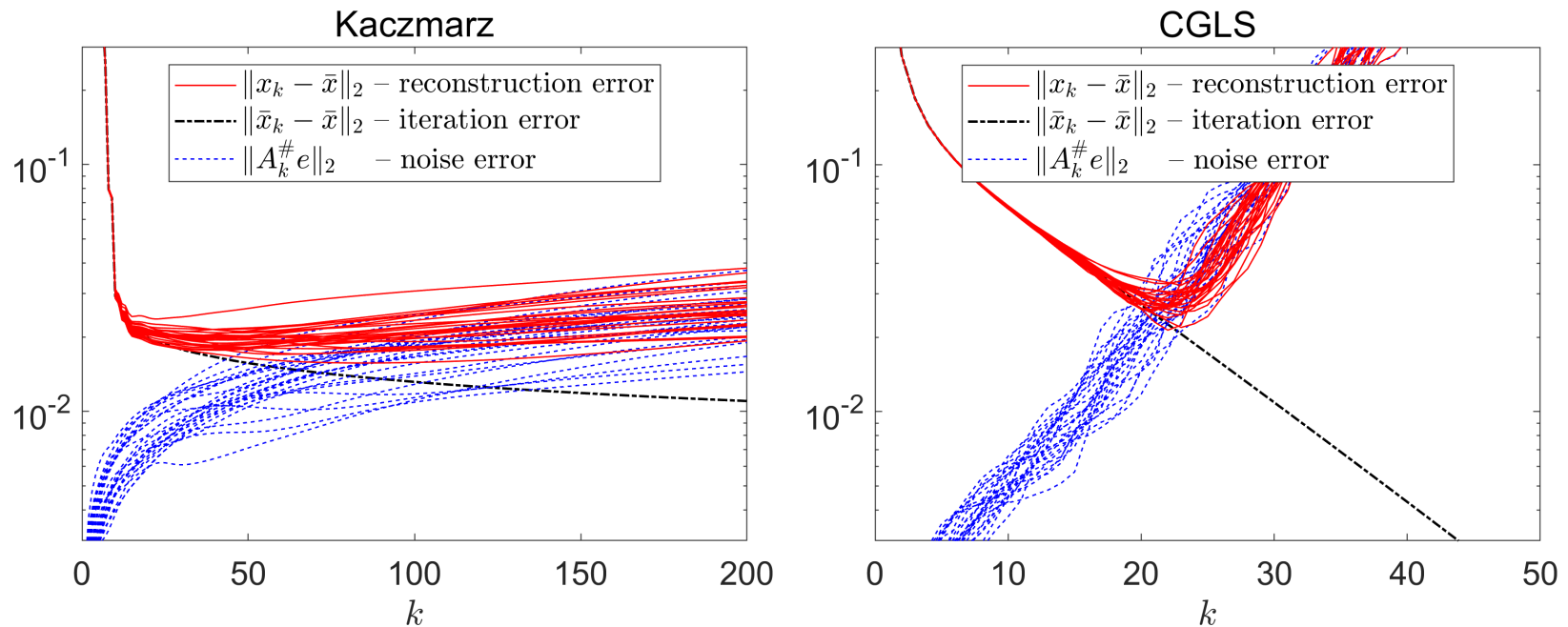


Illustration of the three errors for the gravity test problem, and with 25 realization of the noise.

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:

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

Moreover, if we introduce

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

then we have

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

Then it follows that

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

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

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

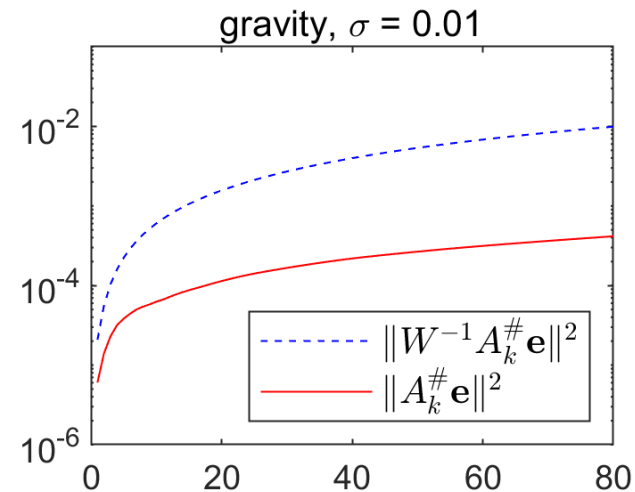
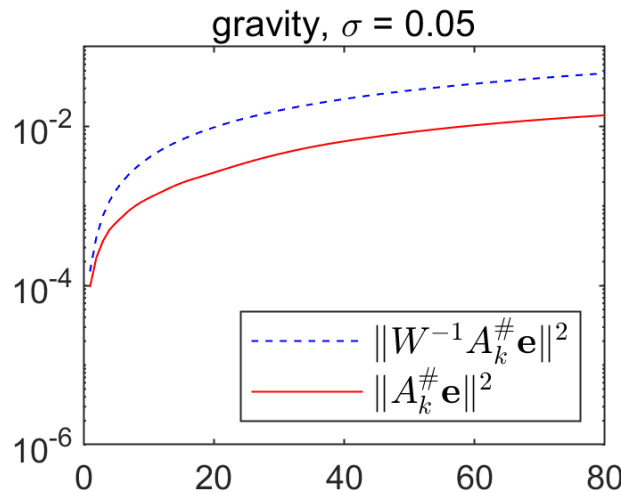
and

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



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

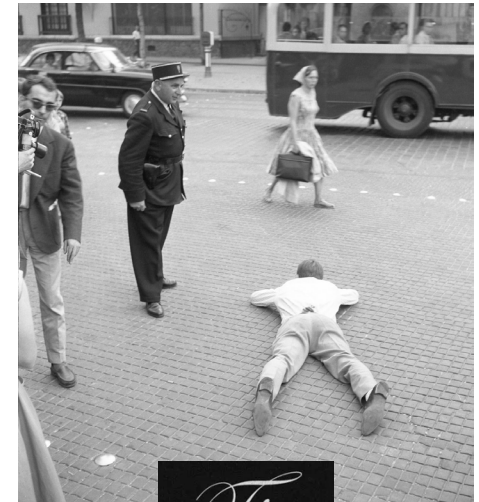
Expected values.



Checking Out

1. Need for analysis of initial convergence.
2. New insight about (near) zero eigenvalues of iteration matrix.
3. Explains fast initial convergence when the solution is rich in the corresponding directions.
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](https://doi.org/10.1007/s10543-025-01098-1).



Fin

Error propagation in Krylov subspace methods is a topic of two newly funded projects → I am looking for a PhD and a postdoc.

Appendix: ART History

Kaczmarz (1937): cyclic sweeps:

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

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

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

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

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

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

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