Tutorial:
Row Action Methods

Per Christian Hansen
Technical University of Denmark

HDtomo
HIGH - DEFINITION TOMOGRAPHY

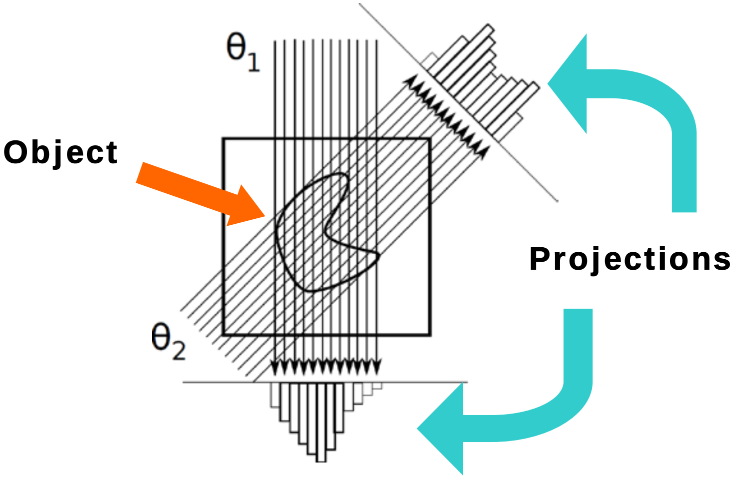
DTU Compute
Department of Applied Mathematics and Computer Science

Topics Covered Today

- Tomography, inverse problems, and direct inversion
- An alternative: the algebraic formulation
- Iterative methods: row action methods (ART & SIRT)
- Convergence and semi-convergence
- Choice of the relaxation parameter
- Stopping rules
- AIR Tools – a new MATLAB® package

What is Tomography?

Image reconstruction from projections



Tomography is the science of *seeing inside objects*.

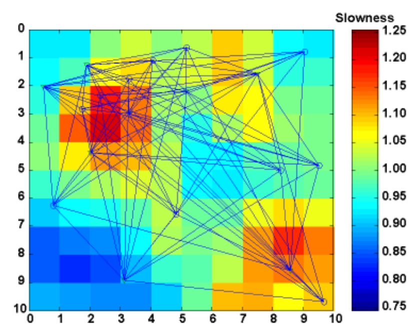
Signals – e.g., waves, particles, currents – are sent through an object from many different angles.

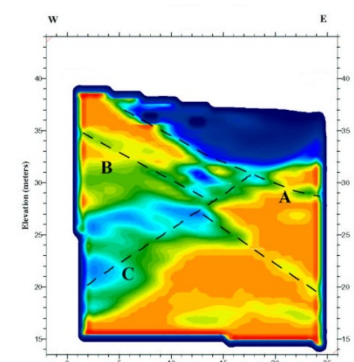
The response of the object to the signal is measured (projections).

We use the data + a mathematical "forward model" to compute a 3D image of the object's interior.

Medical tomography

Seismic tomography





The Origin of Tomography

Johan Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte Längs gewisser Mannings-faltigkeiten*, Berichte Sächsische Akadamie der Wissenschaften, Leipzig, Math.-Phys. Kl., 69, pp. 262-277, 1917.

<u>Main result:</u>

An object can be perfectly reconstructed from a full set of projections.

NOBELFÖRSAMLINGEN KAROLINSKA INSTITUTET

THE NOBEL ASSEMBLY AT THE KAROLINSKA INSTITUTE

11 October 1979

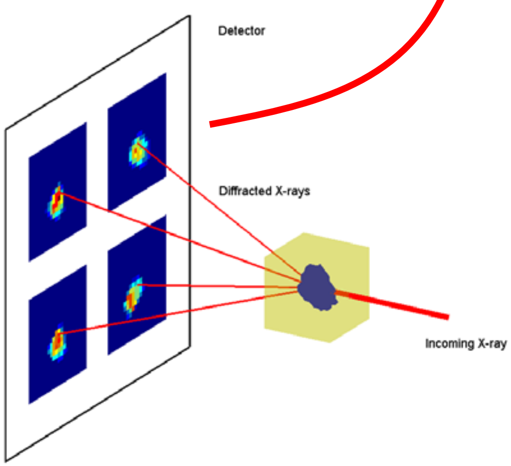
The Nobel Assembly of Karolinska Institutet has decided today to award the Nobel Prize in Physiology or Medicine for 1979 jointly to

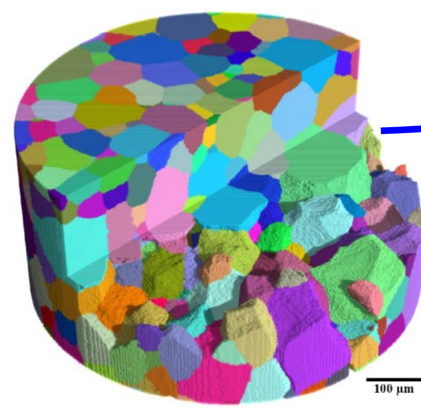
Allan M Cormack and Godfrey Newbold Hounsfield

for the "development of computer assisted tomography".

Inverse Problems

Goal: use measured data to *compute* “hidden” information.





Inverse Problem – Ill-Posed Problem

Hadamard's conditions for a well-posed problem

$$K f = g$$

Existence

The problem must have a solution.

Uniqueness

The must be only one solution to the problem.

Stability

The solution must depend continuously on the data.

The discretized problem

$$A\,x = b$$

→ Reformulation

Replace "=" with $\min \|\cdot\|_2$ → consistent system $A\,x = P_{\mathcal{R}(A)} b$.

→ Add more requirements

Add additional constraint, e.g., minimum 2-norm → $x = A^\dagger\, b$.

→ *Sensitivity* is the problem

$\mathrm{cond}(A) \approx \infty$ → solution x and $\mathrm{rank}(A)$ very sensitive to pert.

Some Reconstruction Algorithms

Direct Inversion Methods

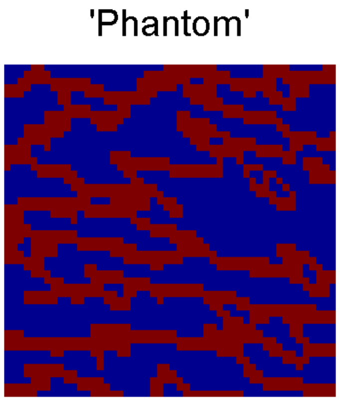
The forward problem is formulated as a certain transform
$\rightarrow$ formulate a stable way to compute the inverse transform.

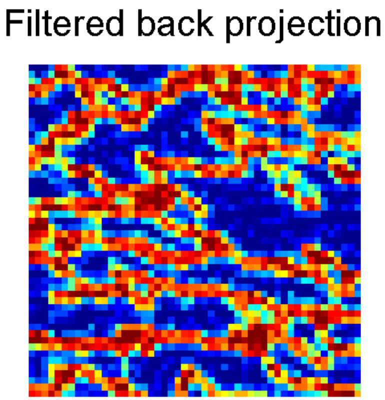
Examples: inverse Radon transform, filtered back projection.

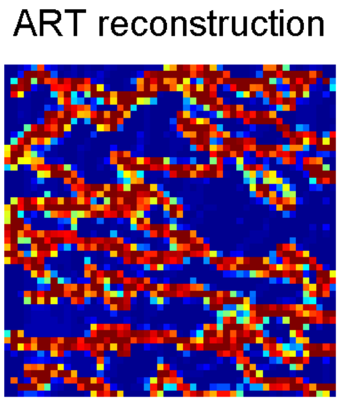
Algebraic Iterative Methods – Row Action Methods

Write the forward problem as an algebraic model $A\,x = b$
$\rightarrow$ reconstruction amounts to solving $A\,x = b$ iteratively.

Examples: ART, Landweber, Cimmino, conjugate gradients.







Filtered Back Projection

The steps of the inverse Radon transform:

Choose a filter: $\mathcal{F}(\omega) = |\omega| \cdot \mathcal{F}_{\mathrm{low-pass}}(\omega)$.

Apply filter for each angle ϕ in the sinogram: $G_\phi(\rho) = \mathrm{ifft}\big(\mathcal{F} \cdot \mathrm{fft}(g_\phi)\big)$.

Back projection to image: $f(x,y) = \int_0^{2\pi} G_\phi(x\cos\phi + y\sin\phi)\, d\phi$.

Interpolation to go from polar to rectanglar coordinates (pixels).

Advantages

- Fast because it relies on FFT computations!
- Low memory requirements.
- Lots of experience with this method from many years of use.

Drawbacks

- Needs many projections for accurate reconstructions.
- Difficult to apply to non-uniform distributions of rays.
- Filtering is “hard wired” into the algorithm (low-pass filter).
- Difficult to incorporate prior information about the solution.

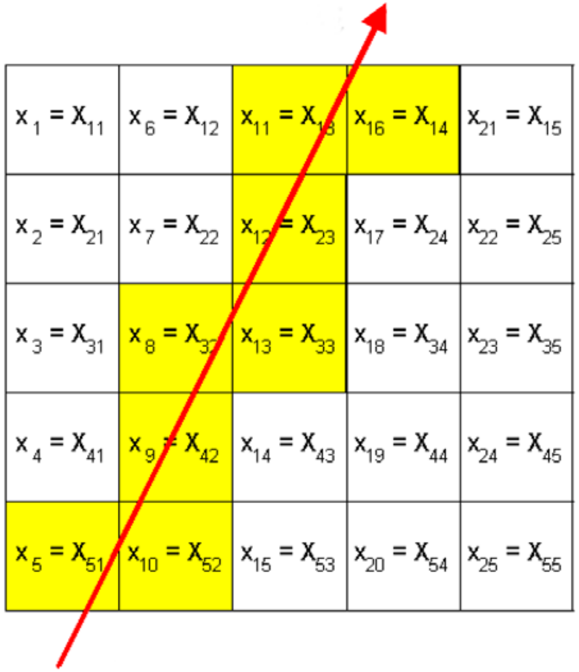
Setting Up the Algebraic Model

Damping of *i*-th X-ray through domain:

$$b_i = \int_{\mathrm{ray}_i} \chi(\mathbf{s})\, d\ell, \qquad \chi(\mathbf{s}) = \text{attenuation coef.}$$

Assume $\chi(\mathbf{s})$ is a constant x_j in pixel *j*, leading to:

$$b_i = \sum_j a_{ij}\, x_j\,, \qquad a_{ij} = \text{length of ray } i \text{ in pixel } j.$$



This leads to a large linear system:

$$A\,x = b \qquad b = b^* + e$$

"Geometry" — Image — Projections — Noise

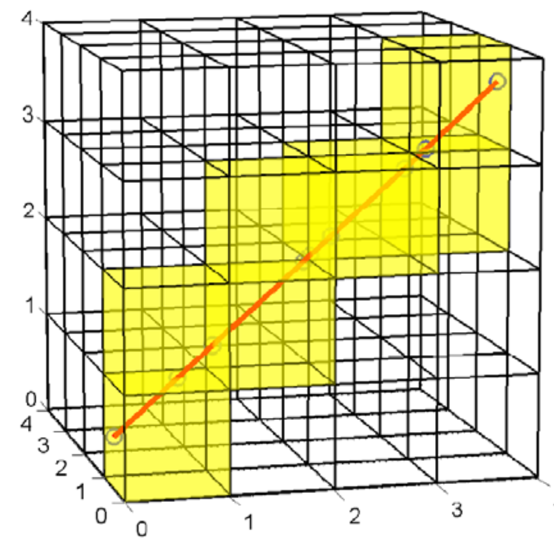
$$b^* = A\,x^*$$

DTU

More About the Coefficient Matrix, 3D Case

$$b_i = \sum_j a_{ij}\, x_j, \qquad a_{ij} = \text{length of ray } i \text{ in voxel } j.$$

To compute the matrix element a_{ij} we simply need to know the intersection of ray i with voxel j. Let ray i be given by the line



$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + t \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \qquad t \in \mathbb{R}.$$

The intersection with the plane $x = p$ is given by

$$\begin{pmatrix} y_j \\ z_j \end{pmatrix} = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} + \frac{p - x_0}{\alpha} \begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \qquad p = 0, 1, 2, \ldots$$

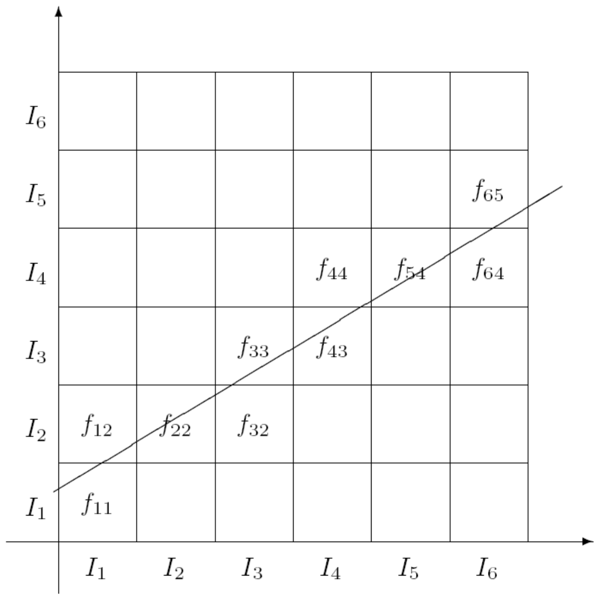
with similar equations for the planes $y = y_j$ and $z = z_j$.

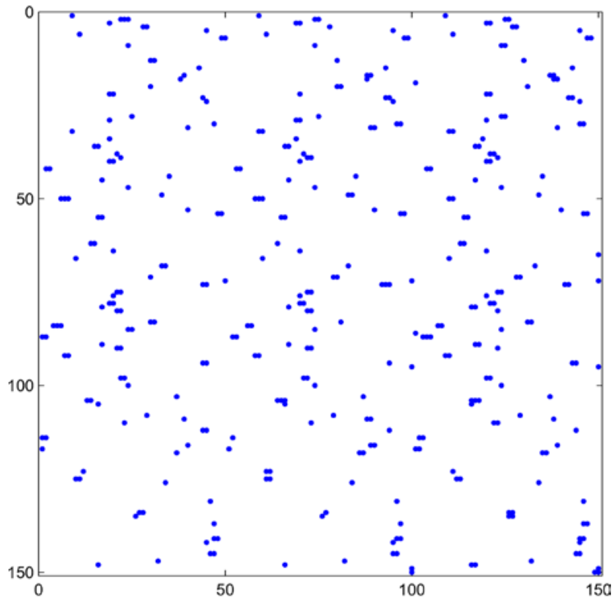
From these intersetions it is easy to compute the ray length in voxel j..

Siddon (1985) presented a fast method for these computations.

The Matrix is Sparse

Each ray intersects only a few cells, hence *A* is very sparse.

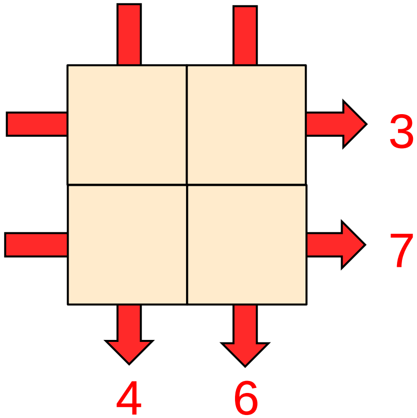




Many rows are structurally orthogonal.

This sparsity plays a role in the convergence and the success of some of the iterative methods.

Analogy: the “Sudoku” Problem – 数独



$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 4 \\ 6 \end{pmatrix}$$

0	3
4	3

1	2
3	4

Infinitely many solutions ($c \in \mathbb{R}$):

$$\begin{bmatrix} & \\ & \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + c \times \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

2	1
2	5

Prior: solution is integer and non-negative →

3	0
1	6

Some Row Action Methods

ART – Algebraic Reconstruction Techniques

- Kaczmarz's method + variants.
- *Sequential* row-action methods that update the solution using one row of A at a time.
- Good semiconvergence observed, but lack of underlying theory of this important phenomenon.

SIRT – Simultaneous Iterative Reconstruction Techniques

- Landweber, Cimmino, CAV, DROP, SART, ...
- These methods use all the rows of A *simultaneously* in one iteration (i.e., they are based on matrix multiplications).
- Slower semiconcergence, but otherwise good understanding of convergence theory.

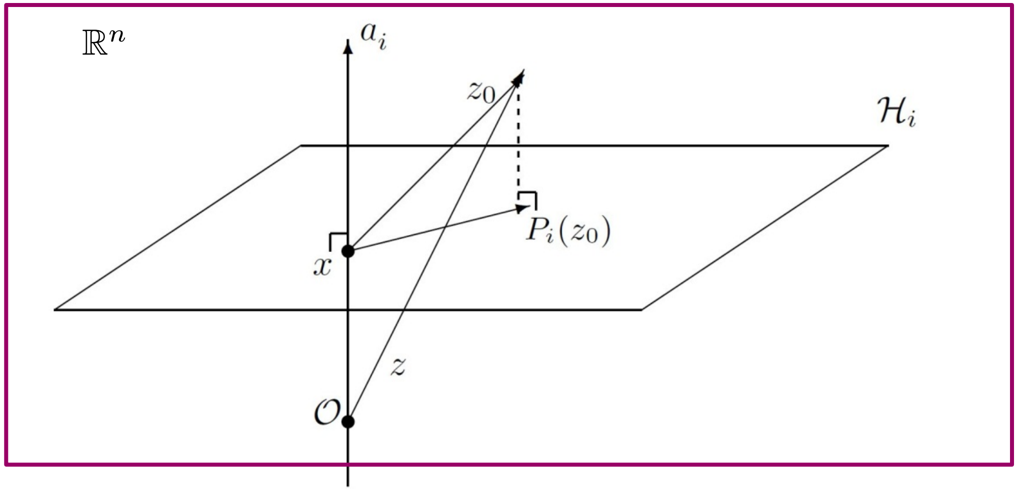
Krylov subspace methods

- CGLS, LSQR, GMRES, ...
- These methods are also based on matrix multiplications

Projection on a Hyperplane

The ith row a_i^T of A and the corresponding element b_i of the rhs b define the hyperplane

$$\mathcal{H}_i = \{x \in \mathbb{R}^n \,|\, a_i^T x = b_i\}.$$

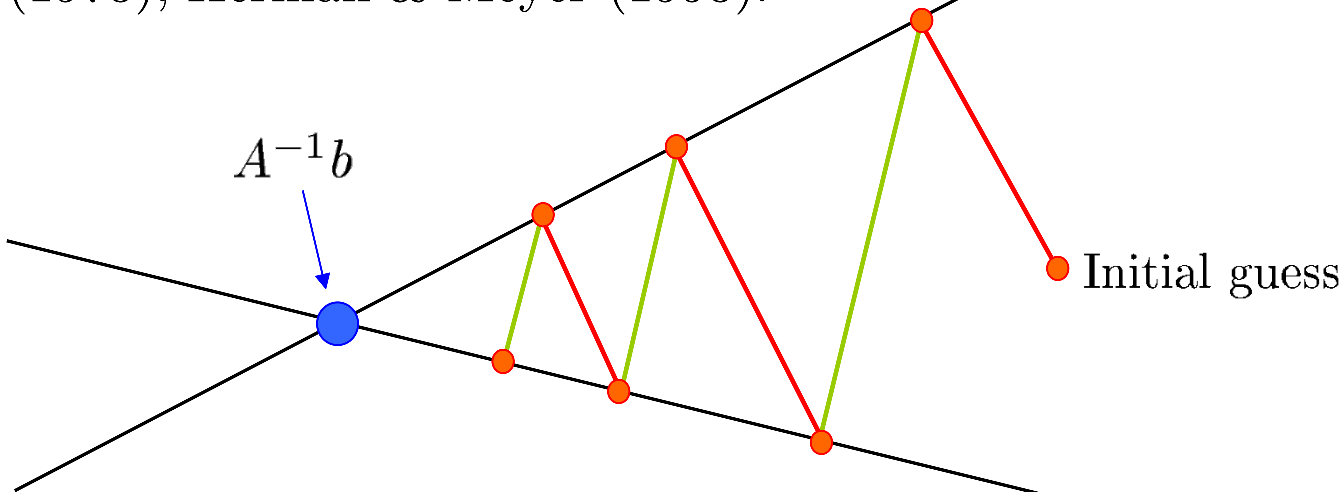


The orthogonal projection of $z \in \mathbb{R}^n$ on $\mathcal{H}_i$ is given by

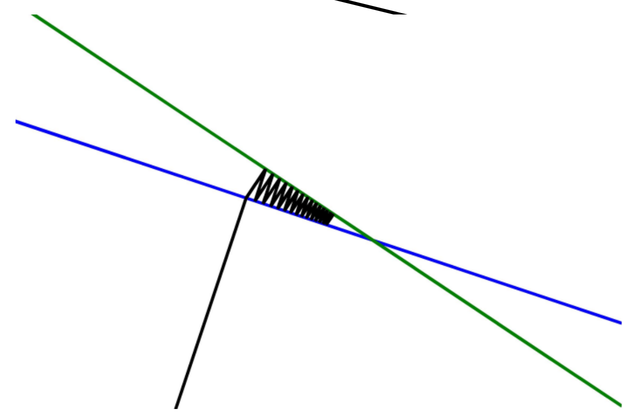
$$P_i(z) = z + \frac{b_i - a_i^T z}{\|a_i\|_2^2} \, a_i.$$

The Geometry of ART

ART has fast initial convergence, and has therefore been a method of choice when only a few iterations can be afforded cf. Gordon, Bender & Herman (1970), Hounsfield (1973), Herman & Meyer (1993).



But after some initial iterations the convergence can be very slow.



Formulation of ART Methods

The typical step in these methods involves the ith row a_i^T of A in the following update of the iteration vector:

$$x \leftarrow x + \lambda \frac{b_i - a_i^T x}{\|a_i\|_2^2} a_i,$$

where λ is a relaxation parameter ($\lambda = 1 \rightarrow$ projection).

Different sweeps of the m rows of A:

Kaczmarz: $i = 1, 2, \ldots, m$.

Symmetric Kaczmarz: $i = 1, 2, \ldots, m-1, m, m-1, \ldots, 3, 2$.

Randomized Kaczmarz: select row i randomly with probability proportional to the row norm $\|a_i\|_2$.

Convergence and Condition Number

Strohmer & Vershynin (2009): Known estimates of convergence rates are based on quantities of A that are hard to compute and difficult to compare with convergence estimates of other iterative methods. What numerical analysts would like to have is estimates of the convergence rate with respect to standard quantities such as $\|A\|$ and $\|A^{-1}\|$. The difficulty: the rate of convergence for ART depends on the ordering of the equations, while $\|A\|$ and $\|A^{-1}\|$ are independent of the ordering.

With random selection of the rows, the expected behavior is:

$$\left(1 - \frac{2k}{\mathrm{cond}(A)^2}\right) \|x^0 - x^*\|_2^2 \;\square$$

$$\mathcal{E}\left(\|x^k - x^*\|_2^2\right) \;\square\; \left(1 - \frac{1}{\mathrm{cond}(A)^2}\right)^k \|x^0 - x^*\|_2^2.$$

Note: $AA^T =$ diagonal matrix $\Rightarrow$ convergence in one sweep!

Towards SIRT – Steepest Descent Method

Consider the least squares problem:

$$\min_x \tfrac{1}{2} \|A\,x - b\|_2^2 \qquad \Leftrightarrow \qquad A^T A\,x = A^T b\ .$$

The gradient for $f(x) = \frac{1}{2}\|A\,x - b\|_2^2$ is $\nabla f(x) = A^T(A\,x - b)$.

The steepest-descent method involves the steps:

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

With projection $\mathcal{P}$ this is the gradient projection algorithm:

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

The SIRT methods are based on this approach. ➡ Next page

SIRT Methods

Diagonally Relaxed Orthogonal Projection

Simultaneous Algebraic Reconstruction Technique

The general form:

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

Some methods use the row norms $\|a_i\|_2$.

Landweber: $T = I$ and $M = I$.

Cimmino: $T = I$ and $M = D = \frac{1}{m} \mathrm{diag}\left(\frac{1}{\|a_i\|_2^2} \right)$.

CAV (component averaging method): $T = I$ and $M = D_S = \mathrm{diag}\left(\frac{1}{\|a_i\|_S^2} \right)$ with $S = \mathrm{diag}(\mathbf{nnz}(\text{column } j))$.

DROP: $T = S^{-1}$ and $M = mD$.

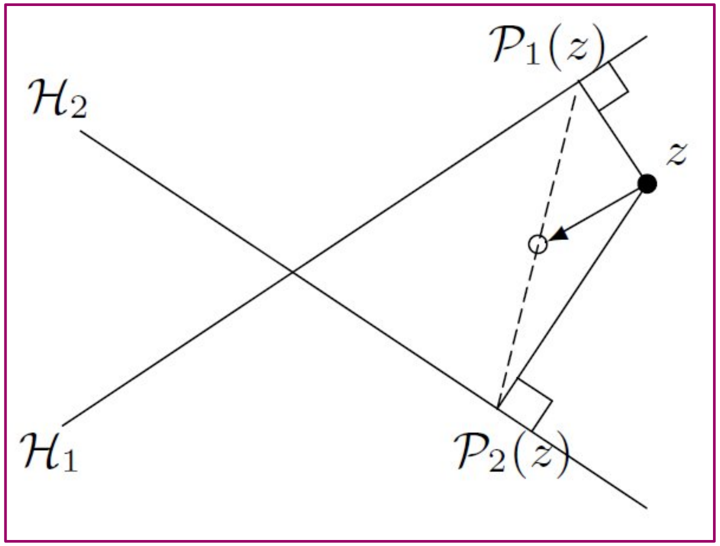
SART: $T = \mathrm{diag}(\text{row sums})^{-1}$ and $M = \mathrm{diag}(\text{column sums})^{-1}$.

The Geometry of Cimmono's Method

Cimmino's method is related to projections on hyperplanes.

Given $z \in \mathbb{R}^n$, the next iterate ($\lambda = 1$) is the average of the projections $P_i(z)$ on the hyperplanes $\mathcal{H}_i$, $i = 1, 2, \ldots, m$.

$$
\begin{aligned}
z^{\text{new}} &= \frac{1}{m} \sum_{i=1}^{m} P_i(z) \\
&= \frac{1}{m} \sum_{i=1}^{m} \left(z + \frac{b_i - a_i^T z}{\|a_i\|_2^2} a_i \right) \\
&= z + \frac{1}{m} \sum_{i=1}^{m} \frac{b_i - a_i^T z}{\|a_i\|_2^2} a_i \\
&= z + A^T D\,(b - A\,z), \qquad D = \frac{1}{m} \text{diag}\left(\frac{1}{\|a_i\|_2^2} \right)
\end{aligned}
$$



Convergence of Landweber

Nesterov (2004): let $\mu = \sigma_{\min}(A)$, then

$$\|x^k - x^*\|_2^2 \;\Box\; \left(1 - \frac{2\lambda\mu^2\|A\|_2^2}{\mu^2 + \|A\|_2^2}\right)^k \|x^0 - x^*\|_2^2 \ .$$

If we choose λ slightly smaller than $2/\|A\|_2^2$ (the largest value that ensures convergence), then

$$\|x^k - x^*\|_2^2 \lesssim \left(\frac{\mathrm{cond}(A)^2 - 1}{\mathrm{cond}(A)^2 + 1}\right)^k \|x^0 - x^*\|_2^2 \ .$$

For general SIRT methods, replace A by $M^{1/2}A$.

Nonnegativity and Box Constraints

It is easy to inporporate a projection $\mathcal{P}$ on a convex set $\mathcal{C}$ in the ART and SIRT iterations:

$$x \leftarrow \mathcal{P}\left(x + \lambda \frac{b_i - a_i^T x}{\|a_i\|_2^2} a_i \right) .$$

$$x \leftarrow \mathcal{P}\big(x + \lambda \, T A^T M \, (b - A\, x) \big) .$$

E.g., $\mathcal{C}$ can represent nonnegativity constraints $(x \geq 0)$ or box constraints (a □ x □ b). Nonneg. in Matlab: `x(x<0) = 0;`

The projected SIRT methods converge to a solution to

$$\min_{x \in \mathcal{C}} \|b - A\, x\|_M \ .$$

Krylov Subspace Methods

In spite of their fast convergence for some problems, these methods are less known in the tomography community.

The most important method is CGLS, obtained by applying the classical Conjugate Gradient method to the least squares problem:

$$x^{(0)} = 0 \quad \text{(starting vector)}$$

$$r^{(0)} = b - A\,x^{(0)}$$

$$d^{(0)} = A^T r^{(0)}$$

for $k = 1, 2, \ldots$

$$\bar{\alpha}_k = \|A^T r^{(k-1)}\|_2^2 / \|A\,d^{(k-1)}\|_2^2$$

$$x^{(k)} = x^{(k-1)} + \bar{\alpha}_k\, d^{(k-1)}$$

$$r^{(k)} = r^{(k-1)} - \bar{\alpha}_k\, A\, d^{(k-1)}$$

$$\bar{\beta}_k = \|A^T r^{(k)}\|_2^2 / \|A^T r^{(k-1)}\|_2^2$$

$$d^{(k)} = A^T r^{(k)} + \bar{\beta}_k\, d^{(k-1)}$$

end

The work:

One mult. with A

One mult. with A^T

Setting the Stage for the Analysis

We need the SVD of the matrix A:

$$A = U\,\Sigma\,V^T = \sum_{i=1}^{\min(m,n)} u_i\,\sigma_i\,v_i^T\,.$$

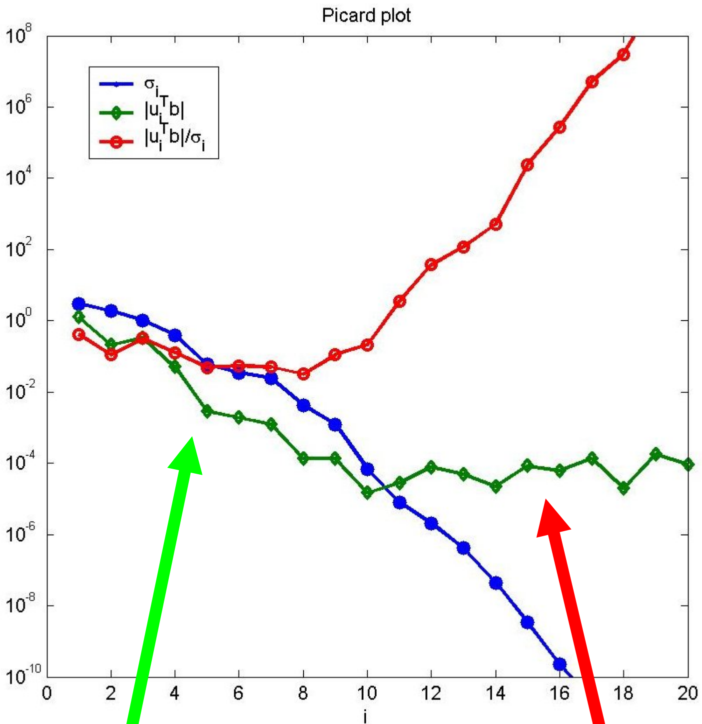
The (minimum norm) least squares least squares solution is:

$$x_{\rm LS} = A^\dagger b = \sum_{i=1}^{\mathrm{rank}(A)} \frac{u_i^T b}{\sigma_i}\,v_i\,.$$

Regularized solutions (obtained by "spectral filtering") are:

$$x_{\rm reg} = \sum_{i=1}^{n} \varphi_i\,\frac{u_i^T b}{\sigma_i}\,v_i\,, \qquad \varphi_i = \text{filter factors.}$$

The Need for Regularization



Assume Gaussian noise:

$$b = b^* + e \,, \quad e \sim \mathcal{N}\left(0, \sigma_{\text{noise}}^2 I\right).$$

Then

$$x_{\text{naive}} \equiv A^{-1} b = x^* + A^{-1} e,$$

and using the SVD we see that

$$x_{\text{naive}} = \sum_{i=1}^n \frac{u_i^T b}{\sigma_i} v_i$$

$$= \sum_{i=1}^n \frac{u_i^T b^*}{\sigma_i} v_i + \sum_{i=1}^n \frac{u_i^T e}{\sigma_i} v_i.$$

"inverted noise"

<u>Picard condition:</u>

$|u_i^T b|$ decays faster than σ_i for small i.

<u>Noise:</u>

$|u_i^T b|$ levels off for larger i.

Regularization:

keep the "good" SVD components and discard the noisy ones!

Convergence of the Iterative Methods

Assume that the solution is *smooth*, as controlled by a parameter $\alpha > 0$,

$$u_i^T b = \sigma_i^{1+\alpha}, \qquad i = 1, \ldots, n,$$

and that the right-hand side has no errors/noise.

Then the iterates x^k converge to an exact solution $x^* \in \mathcal{R}(A^T)$ as follows.

ART and SIRT methods:

$$\|x^k - x^*\|_2 = \mathcal{O}(k^{-\alpha/2}), \qquad k = 0, 1, 2, \ldots$$

CGLS:

$$\|x^k - x^*\|_2 = \mathcal{O}(k^{-\alpha}), \qquad k = 0, 1, 2, \ldots$$

The interesting case is when errors/noise is present in the right-hand side!

Enter the Noise!

In principle, we want to solve $\boxed{A\,x = b}$ – but A is very ill conditioned!

The underlying *noise model:*

$$b = A\,x^* + e, \qquad x^* = \text{exact solution} \qquad e = \text{noise}.$$

It follows that

$$x_{\text{naive}} = A^{-1}b = x^* + A^{-1}e \qquad \text{where} \qquad \|A^{-1}e\|_2 \gg \|x_{\text{exact}}\|_2.$$

Hence the "naive solution" $A^{-1}b$ is useless.

The goal of regularization:

Find an x such that $\|A\,x - b\|_2$ is small (good fit) and x resembles x^*.

Semi-Convergence of the Iterative Methods

Noise model: $b = A\,x^* + e$, where $x^* =$ exact solution, $e =$ additive noise.

Throughout all the iterations, the residual norm $\|A\,x^k - b\|_2$ decreases as the iterates x^k converge to the least squares solution x_{LS}.

But x_{LS} is dominated by errors from the noisy right-hand side b!

However, during the first iterations, the iterates x^k capture "important" information in b, associated with the exact data $b^* = A\,x^*$.

- In this phase, the iterates x^k approach the exact solution x^*.

At later stages, the iterates starts to capture undesired noise components.

- Now the iterates x^k diverge from the exact solution and they approach the undesiredleast squares solution x_{LS}.

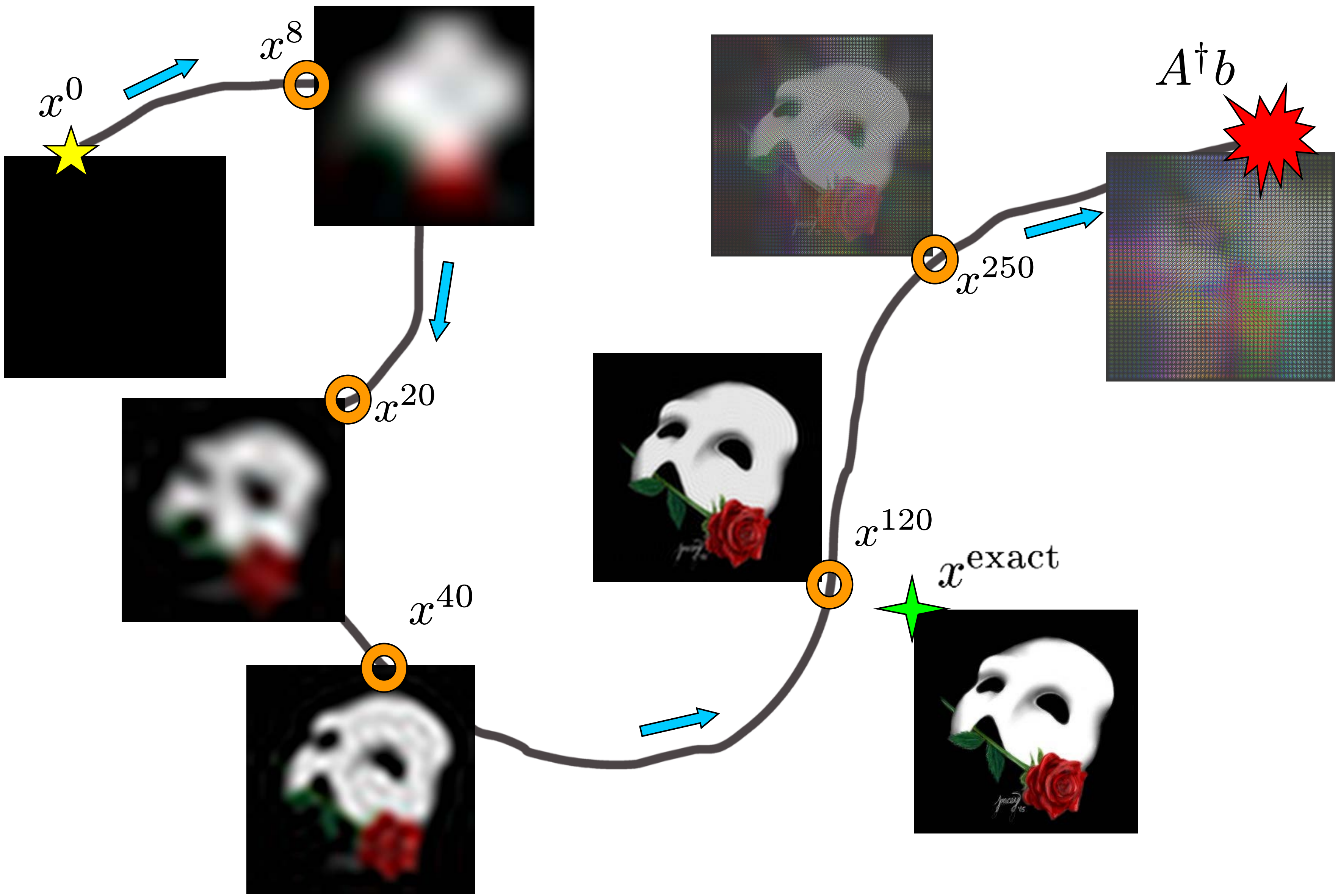
This behavior is called *semi-convergence*, a term coined by Natterer (1986).

"... even if [the iterative method] provides a satisfactory solution after a certain number of iterations, it deteriorates if the iteration goes on."

Many Studies of Semi-Convergence

- G. Nolet, *Solving or resolving inadequate and noisy tomographic systems* (1985)
- A. S. Nemirovskii, *The regularizing properties of the adjoint gradient method in ill-posed problems* (1986)
- F. Natterer, *The Mathematics of Computerized Tomography* (1986)
- Brakhage, *On ill-posed problems and the method of conjugate gradients* (1987).
- C. R. Vogel, *Solving ill-conditioned linear systems using the conjugate gradient method* (1887)
- A. van der Sluis & H. van der Vorst, *SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems* (1990)
- M. Hanke, *Accelerated Landweber iterations for the solution of ill-posed equations* (1991).
- M. Bertero & P. Boccacci, *Inverse Problems in Imaging* (1998)
- M. Kilmer & G. W. Stewart, *Iterative regularization and MINRES* (1999)
- H. W. Engl, M. Hanke & A. Neubauer, *Regularization of Inverse Problems* (2000)

Illustration of Semi-Convergence

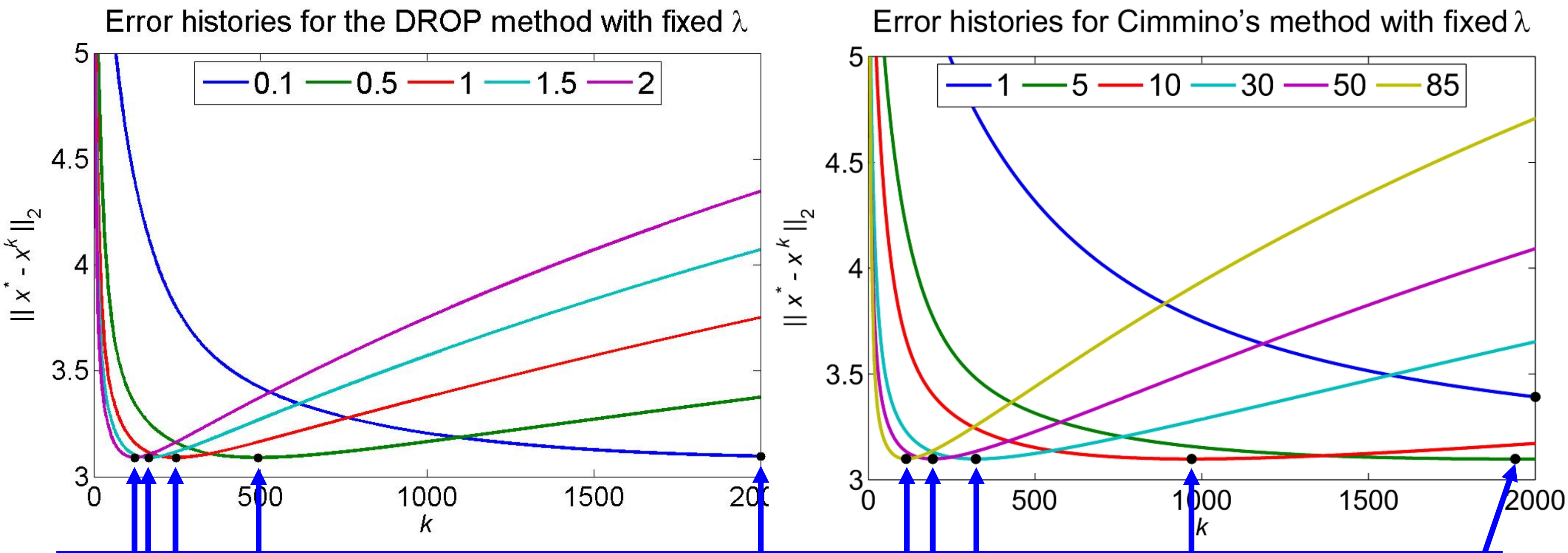


Another Look at Semi-Convergence

Notation: $b = A\,x^* + e$, x^* = exact solution, e = noise.

Initial iterations: the error $\|x^* - x^k\|_2$ decreases.

Later: the error increases as $x^k \rightarrow \text{argmin}_x \|A\,x - b\|_M$.



The minimum error is *independent* of both λ and the method.

Analysis of Semi-Convergence, Fixed λ

Consider the SIRT methods with $T = I$ and the SVD:

$$M^{1/2} A = U \Sigma V^T = \sum_{i=1}^n u_i \, \sigma_i \, v_i^T ,$$

Then x^k is a filtered SVD solution:

$$x^k = \sum_{i=1}^n \varphi_i^{[k]} \frac{u_i^T M^{1/2} b}{\sigma_i} v_i , \qquad \varphi_i^{[k]} = 1 - \left(1 - \lambda \, \sigma_i^2\right)^k .$$

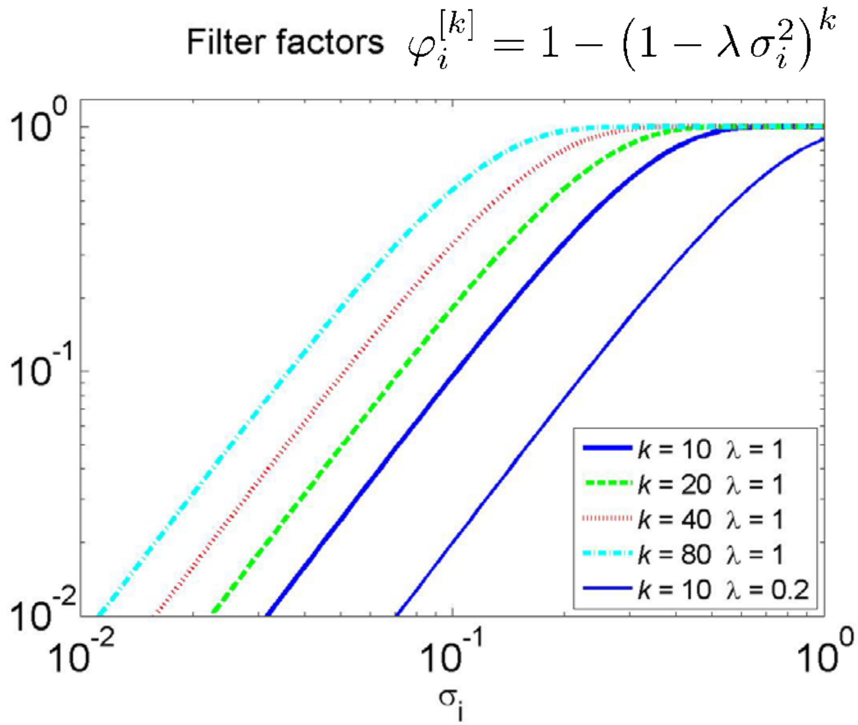
The ith component of the error, in the SVD basis, is

$$v_i^T (x^* - x^k) = (1 - \varphi_i^{[k]}) \, v_i^T x^* - \varphi_i^{[k]} \frac{u_i^T M^{1/2} e}{\sigma_i} .$$

IE_i^k: iteration error

NE_i^k: noise error

The Behavior of the Filter Factors



When k doubles, the "break point" is reduced by a factor $\approx \sqrt{2}$.

The filter factors *dampen* the "inverted noise" $u_i^T M^{1/2} e / \sigma_i$.

$$\lambda\,\sigma_i^2 \ll 1 \Rightarrow \varphi_i^{[k]} \approx k\,\lambda\,\sigma_i^2 \Rightarrow k \text{ and } \lambda \text{ play the same role.}$$

Iteration error $\mathsf{IE}_i^k = (1 - \lambda\sigma_i^2)^k\, v_i^T x^* \to 0$ for $k \to 0$.

About the Noise Error

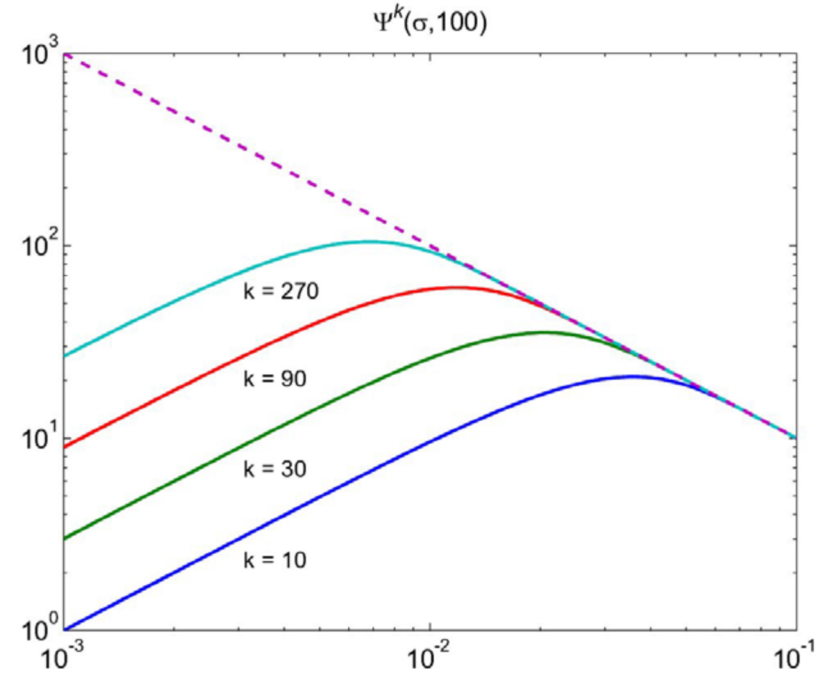
$$\mathsf{NE}_i^k = \Psi^k(\sigma_i, \lambda)\, u_i^T M^{1/2} e$$

$$\Psi^k(\sigma, \lambda) = \frac{1 - (1 - \lambda\, \sigma^2)^k}{\sigma}$$

Fix σ and λ: $\Psi^k \nearrow$ with k.

Fix λ:
max of Ψ^k is attained for

$$\sigma = \sigma_k^* = \sqrt{\frac{1-\zeta_k}{\lambda}}$$



where ζ_k is the unique root in $(0, 1)$ of

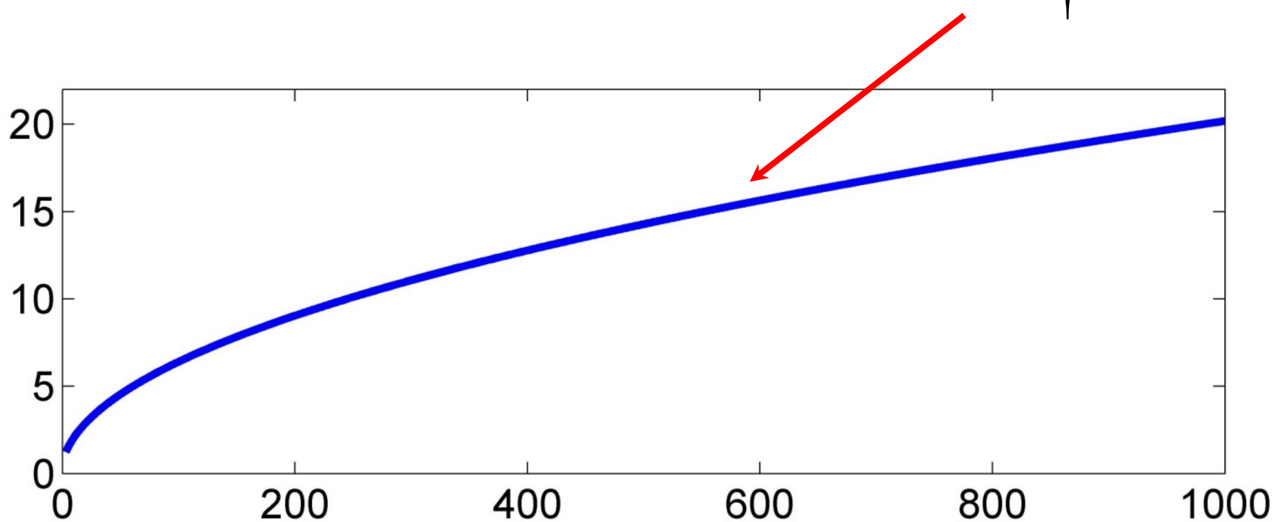
$$g_{k-1}(y) = (2k-1)y^{k-1} - (y^{k-2} + \ldots + y + 1) = 0.$$

Bounding the Noise Error

Assuming a *fixed* λ we thus have

$$\|\mathsf{NE}^k\|_2 \quad \square \quad \max_i \Psi^k(\sigma_i, \lambda)\, \|M^{1/2} e\|_2$$

$$\square \quad \Psi^k(\sigma_k^*, \lambda)\, \|M^{1/2} e\|_2 = \sqrt{\lambda}\, \underbrace{\frac{1-\zeta_k^k}{\sqrt{1-\zeta_k}}}\, \|M^{1/2} e\|_2 .$$



Semi-Convergence of *Projected* SIRT

When the projection $\mathcal{P}$ is included it is not possible the perform a componentwise analysis.

Assuming a *fixed* λ we have shown that

$$\|\mathsf{IE}^k\|_2 \quad \square \quad (1 - \lambda\,\sigma_n^2)^k\,\|x^*\|_2,$$

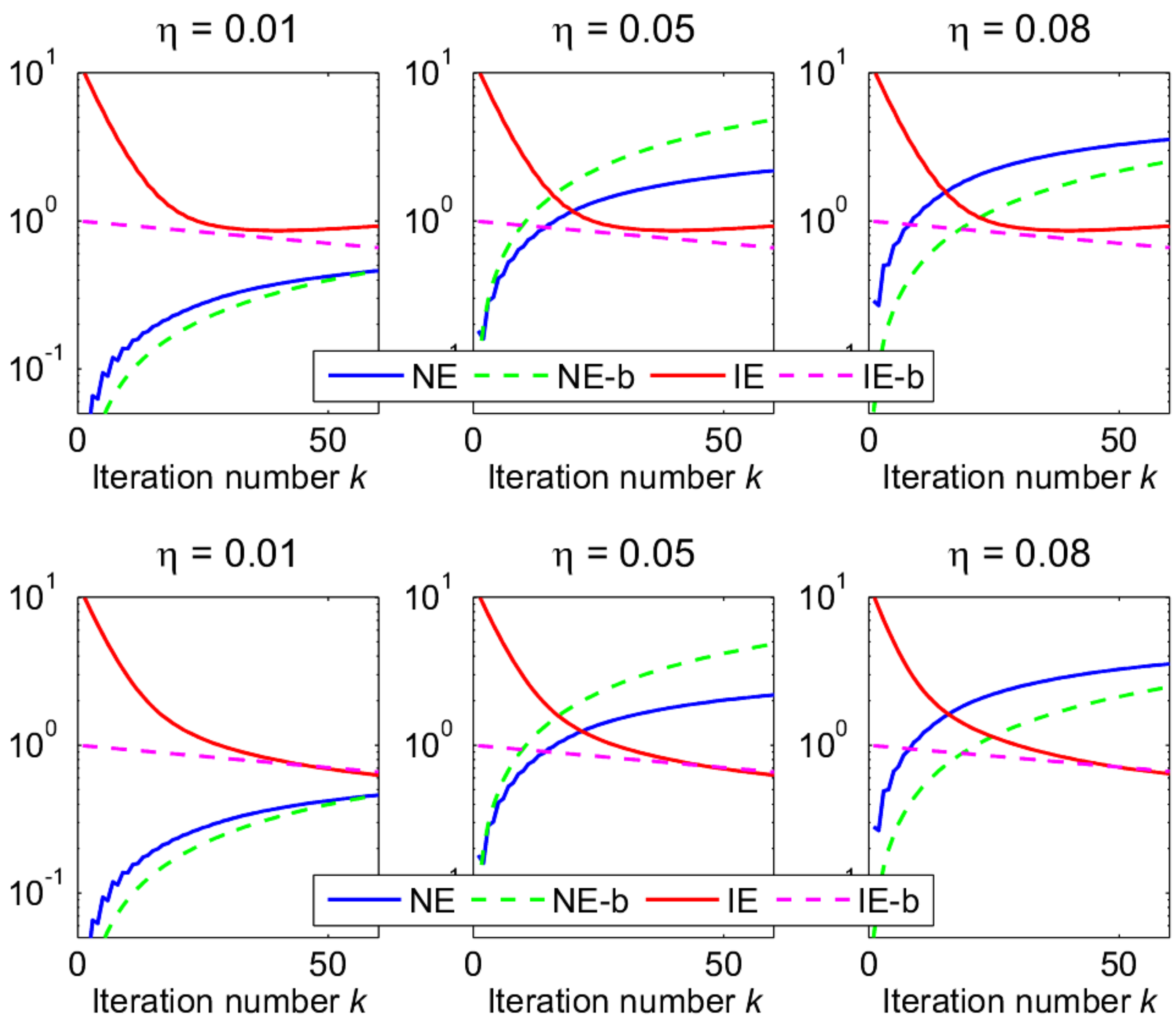
$$\|\mathsf{NE}^k\|_2 \quad \square \quad \mathrm{cond}(A)\,\Psi^k(\sigma_n,\lambda)\,\|M^{1/2}e\|_2.$$

These bounds are very pessimistic – but they correctly *track* the behavior of the iteration and noise error. → Next page.

For $\lambda\,\sigma_n^2 \ll 1$ we have (similar to the unprojected case):

$$\|\mathsf{NE}^k\|_2 \approx \lambda\,k\,\sigma_1\,\|M^{1/2}e\|_2.$$

Iteration and Noise Error, Projected SIRT



Top: consistent system $A\,x = b$.

Bottom: inconsistent system $A\,x \approx b$.

NE-b and IE-b: the pesimistic factors $\mathrm{cond}(A)$ and $\|x^*\|_2$ are omitted.

Analysis of Semi-Convergence – ART

Not much theory has been developed for the semi-convergence of ART.

A first attept:

T. Elfving, P. C. Hansen, and T. Nikazad, *Semi-convergence properties of Kaczmarzs method*, Inverse Problems, 30 (2014):

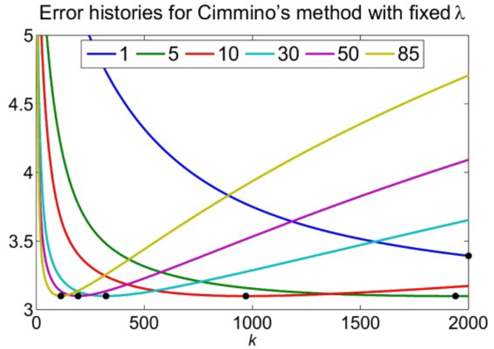
$$\|\text{noise-error}_k\|_2 \; \square \; \frac{\sqrt{\lambda}\delta}{\sigma_{\min}}\sqrt{k} + \mathcal{O}(\sigma_{\min}^2).$$

Choosing the SIRT Relaxation Parameter

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

Goal: fast semi-convergence to the minimum error.

Training. Using a noisy test problem, find the *fixed* $\lambda_k = \lambda$ that gives fastest semi-convergence to the minimum error. Algorithm available in AIR Tools.



Line search (Dos Santos, Appleby & Smolarski, Dax). Minimize the error $\|x^k - x^*\|_2$ in each iteration – must assume that $A\,x = b$ is consistent. When $T = I$ we get:

$$\lambda_k = (r^k)^T M r^k \,/\, \|A^T M r^k\|_2^2, \qquad r^k = b - A\,x^k.$$

A New Strategy: Limit the Noise Error

Assume we used a *fixed* λ in steps $1, \ldots, k-1$; then

$$\begin{aligned} \|\mathsf{NE}^k\|_2 &\leq \max_i \Psi^k(\sigma_i, \lambda)\, \|M^{1/2} e\|_2 \\ &\leq \Psi^k(\sigma_k^*, \lambda)\, \|M^{1/2} e\|_2 = \sqrt{\lambda}\, \frac{1-\zeta_k^k}{\sqrt{1-\zeta_k}}\, \|M^{1/2} e\|_2 . \end{aligned}$$

<u>Strategy Ψ_1:</u> choose $\lambda_0 = \lambda_1 = \sqrt{2}/\sigma_1^2$ and

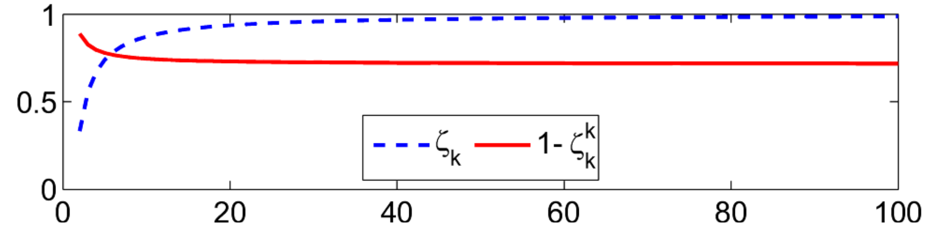
$$\lambda_k = \frac{2}{\sigma_1^2}\,(1-\zeta_k), \qquad k = 2, 3, \ldots$$

<u>Strategy Ψ_2:</u> choose $\lambda_0 = \lambda_1 = \sqrt{2}/\sigma_1^2$ and

$$\lambda_k = \frac{2}{\sigma_1^2}\,\frac{1-\zeta_k}{(1-\zeta_k^k)^2}, \qquad k = 2, 3, \ldots$$

Our New Strategy: What We Achieve

For both variants we obtain relaxation parameters $\lambda_k > 0$ that lead to a *diminishing step-size strategy* with $\lambda_k \to 0$ such that $\sum_k \lambda_k = \infty$.



As a result:

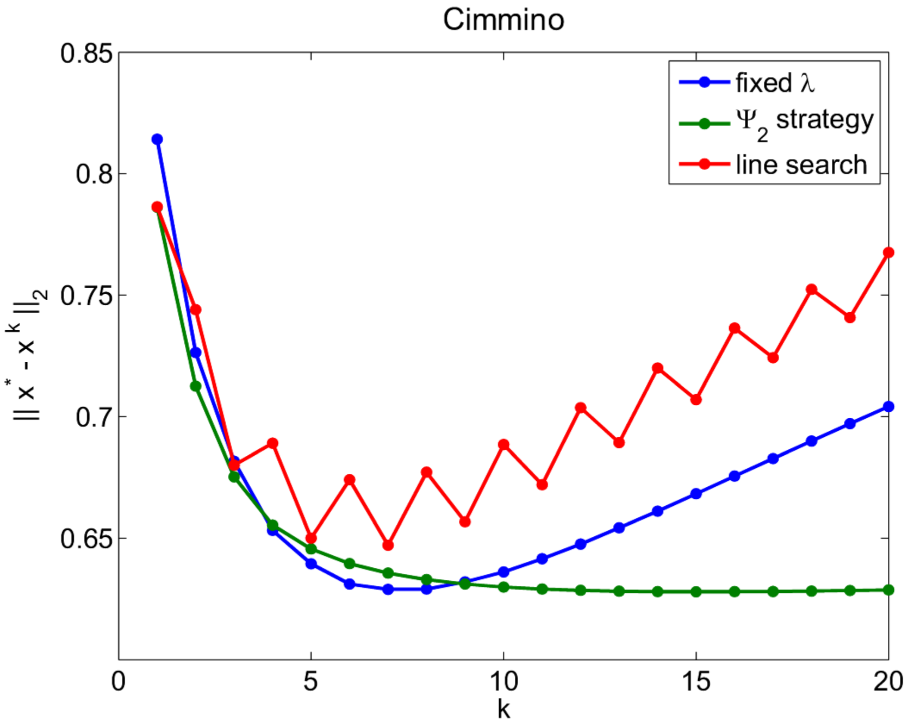
$$\|\mathsf{NE}^k\|_2 \lesssim \frac{\sqrt{2}}{\sigma_1}\,(1-\zeta_k^k)\,\|M^{1/2}e\|_2 \quad \text{for strategy } \Psi_1$$

$$\|\mathsf{NE}^k\|_2 \lesssim \frac{\sqrt{2}}{\sigma_1}\,\|M^{1/2}e\|_2 \quad \text{for strategy } \Psi_2$$

Also, for both variants we still have convergence:

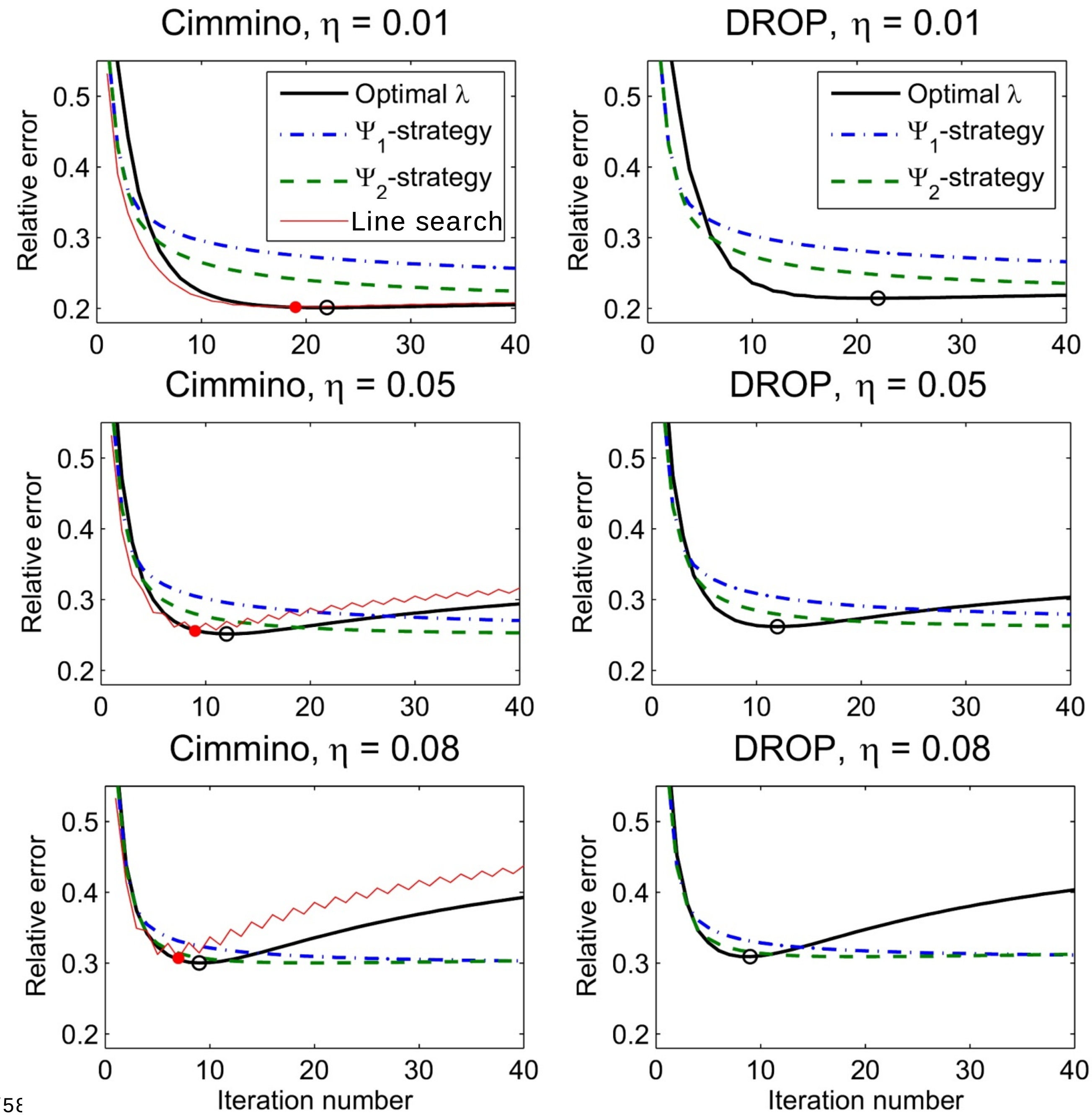
$$x^k \to \mathrm{argmin}\|A\,x - b\|_M \quad \text{as} \quad k \to \infty.$$

Error Histories for Cimmino Example



All three strategies give fast semi-convergence:

- The fixed λ requires training and thus a realistic test problem.
- The line search strategy often gives a ‘zig-zag’ behavior.
- Our new strategy clearly controls the noise propagation.



For high noise levels η, our new strategies "track" the optimal λ.

Line search strategy has zig-zag behavior.

The *same* behavior is observed for the projected methods!

Stopping the Iterations

For iterative methods it is common to stop the iterations when the residul norm $\|b - A\,x^k\|_2$ is "sufficiently small" since this may imply that x^k is close to the solution $A^{-1}b$.

For discritizations of inverse problems, this is problematic:

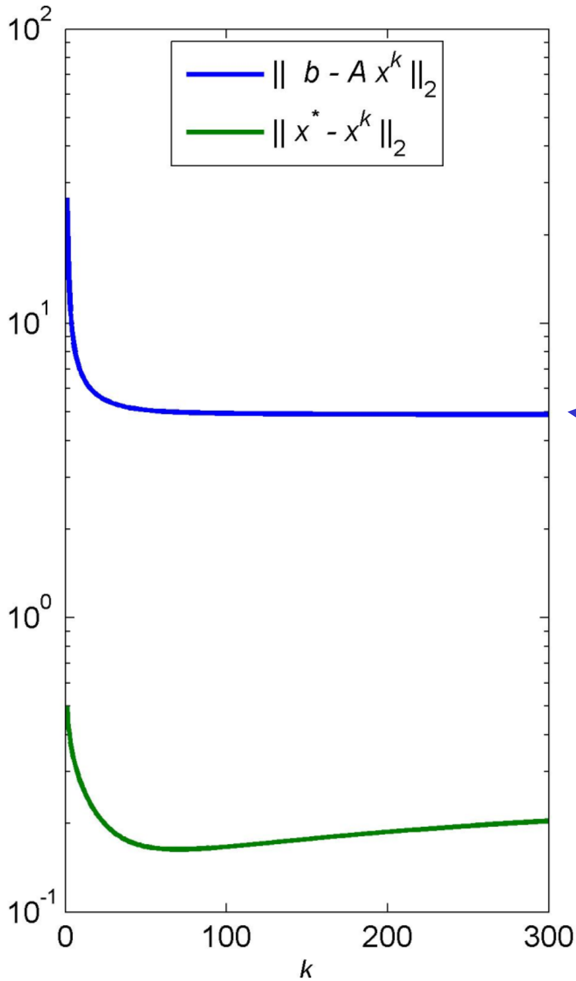
- For ill-conditioned problems a small residual does not imply an accurate solution, since

$$\frac{\|x_{\text{naive}} - x^k\|_2}{\|x_{\text{naive}}\|_2} \;\Box\; \text{cond}(A)\,\frac{\|b - A\,x^k\|_2}{\|b\|_2}\;.$$

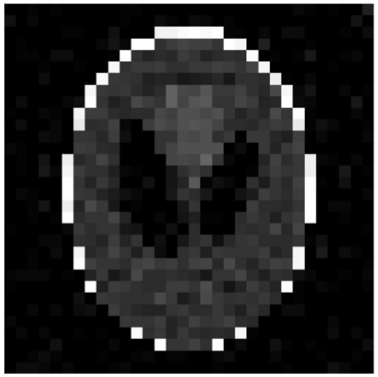
- We do not want to compute $x_{\text{naive}} = A^{-1}b$ in the first place.

Next page

Typical Behavior of Residual Norm and Error



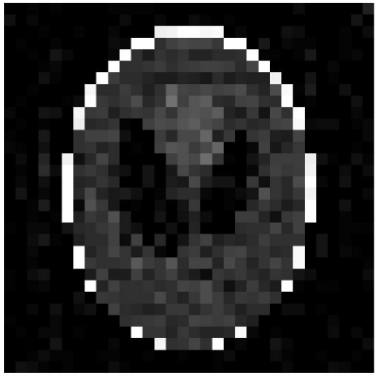
$k = 70$



"Best" reconstruction

The residual norm keeps decreasing

$k = 300$



Stopping Rules for Inverse Problems

Let $\delta = \|e\|_2$, $\quad \tau =$ fudge parameter found by training, and $r_M^k = M^{1/2}(b - A\,x^k)$. Find the smallest k such that:

Discrepancy principle:

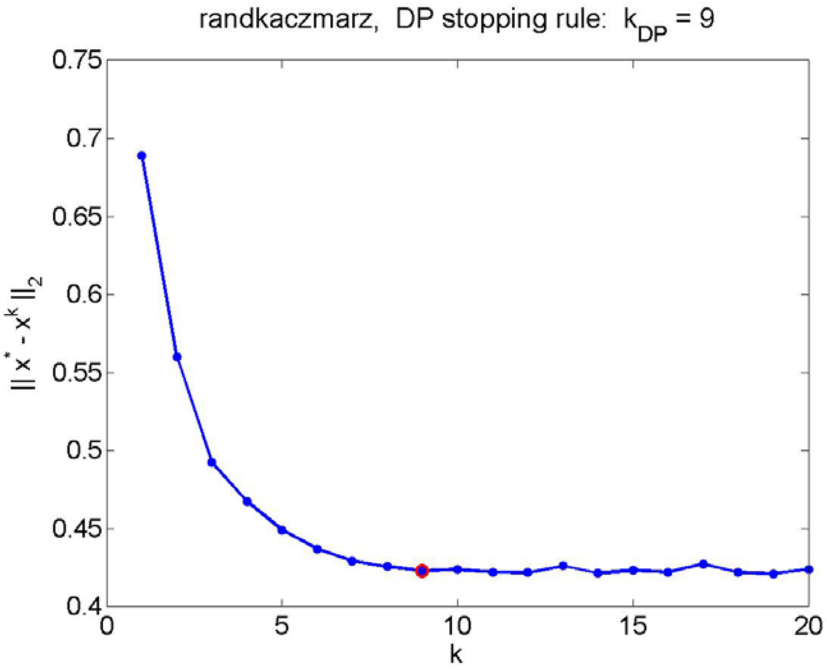
$$\begin{cases} \|r_M^k\|_2 \le \tau\,\delta\,\|M^{1/2}\|_2 & \text{SIRT methods with } T = I \\ \|r^k\|_2 \le \tau\,\delta & \text{all other methods.} \end{cases}$$

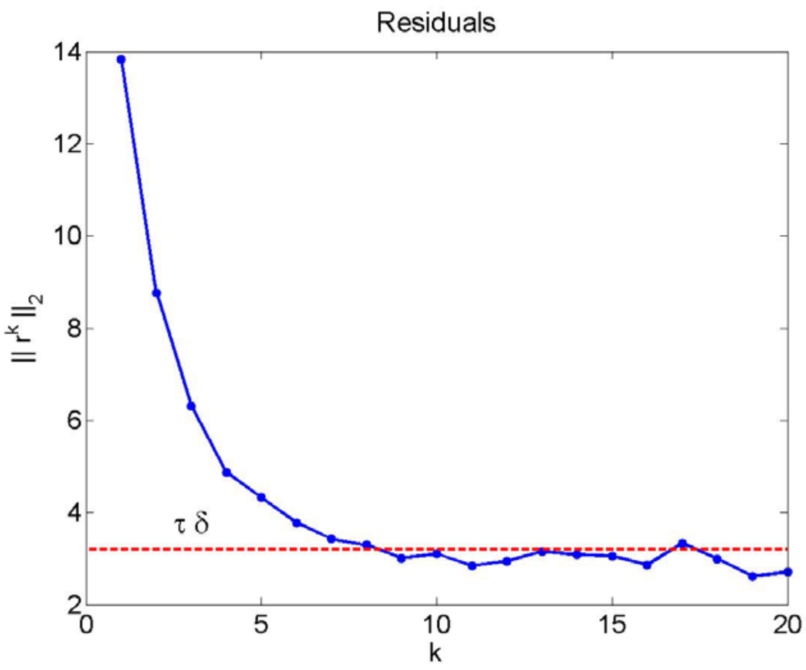
Monotone error rule (SIRT methods only):

$$\frac{(r_M^k)^T(r_M^k + r_M^{k+1})}{\|r_M^k\|_2} \le \tau\,\delta\,\|M^{1/2}\|_2.$$

NCP = normalized cumulative periodogram (Bert Rust): stop when the residual can be considered as noise.

Re. the Discrepancy Principle





For some methods the residual norms do not decay monotonically.

We stop when the residual norm is below $\tau\delta$ for the first time.

NCP = Normalized Cum. Periodogram

The NCP measures the frequency contents in a signal $s \in \mathbb{R}^n$.

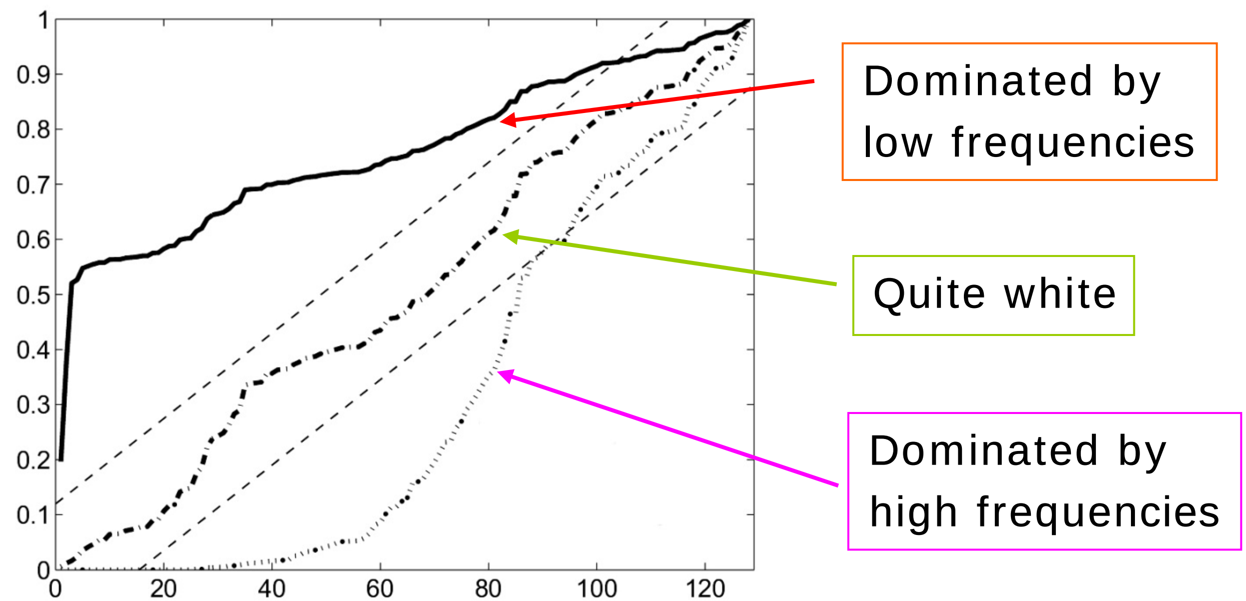
Let $\hat{s} = \mathrm{dft}(s)$ (Fourier transform of s) and $p = |\hat{s}|^2$ (power spectrum of s).

The NCP is a plot of the vector c with elements (assume Matlab indexing):

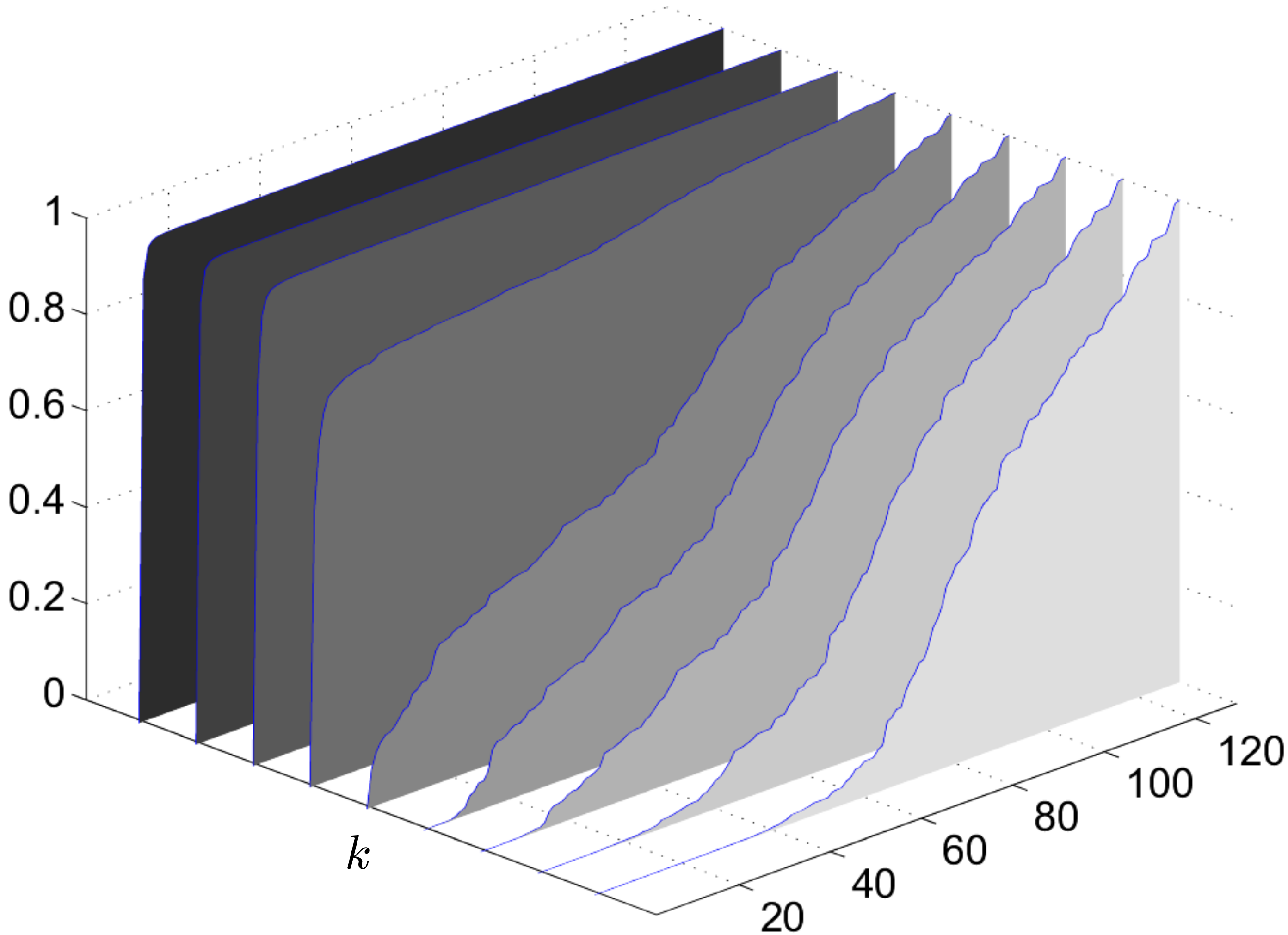
$$c_\ell = \frac{\sum_{i=1}^{\ell} p(i+1)}{\sum_{i=1}^{q} p(i+1)} = \frac{\|\hat{s}(2{:}\,\ell+1)\|_2^2}{\|\hat{s}(2{:}\,q+1)\|_2^2}, \qquad \ell = 1, 2, \ldots, q, \quad q = \lfloor n/2 \rfloor$$

The closer c is to a straight line, the more white the signal s.

Examples with $n = 256$ and $q = 128$.



NCPs for `deriv2` Test Problem



AIR Tools – A MATLAB Package of Algebraic Iterative Reconstruction Methods

- Some important algebraic iterative reconstruction methods
- presented in a common framework
- using identical functions calls,
- and with easy access to:
  - strategies for choosing the relaxation parameter,
  - strategies for stopping the iterations.

> The package allows the user to easily test and compare different methods and strategies on test problems.

Also: "model implementations" for dedicated software (Fortran, C, Python, ...).

Contents of the Package

ART – Algebraic Reconstruction Techniques

- Kaczmarz's method + symmetric and randomized variants.
- Row-action methods that treat one row of A at a time.

SIRT – Simultaneous Iterative Reconstruction Techniques

- Landweber, Cimmino, CAV, DROP, SART.
- These methods are based on matrix multiplications.

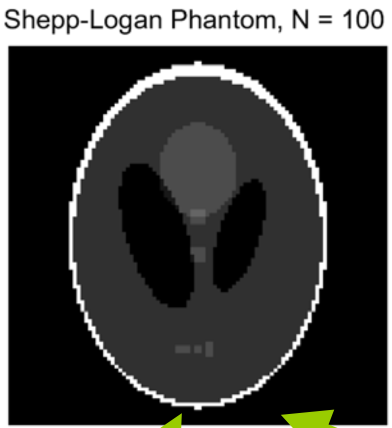
Making the methods useful

- Choice of relaxation parameter λ.
- Stopping rules for semi-convergence.
- Non-negativity constraints.

Tomography test problems

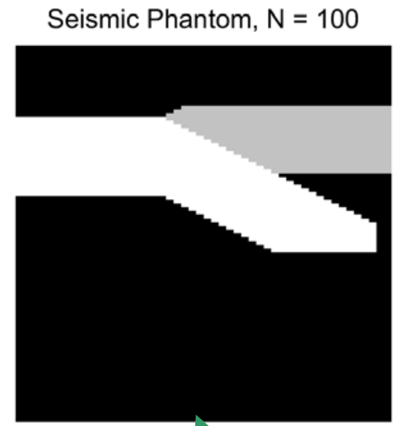
- Medical X-ray (parallel beam, fan beam), seismic travel-time, binary and smooth images (parallel beam)

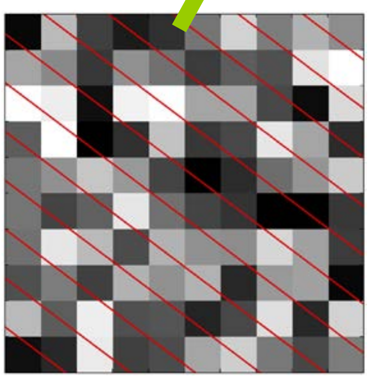
Tomography Test Problems



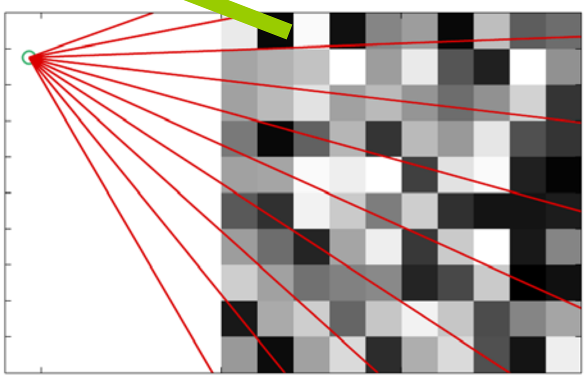
Medical X-ray

Seismic travel time

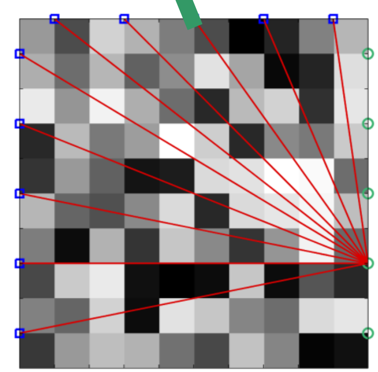




Parallel beam



Fan beam



Sources and receives

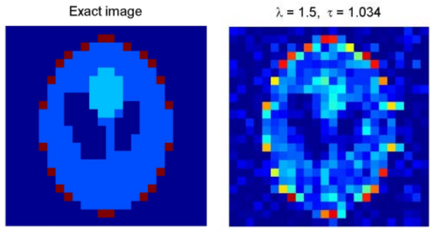
Better medical test problems: use the SNARK09 software from CUNY.

Using AIR Tools – An Example

```
N = 24;          % Problem size is N-by-N.
eta = 0.05;      % Relative noise level.
kmax = 20;       % Number of of iterations.

[A,bex,xex] = fanbeamtomo(N,10:10:180,32); % Test problem
nx = norm(xex); e = randn(size(bex));      %  with noise.
e = eta*norm(bex)*e/norm(e); b = bex + e;

lambda = trainLambdaSIRT(A,b,xex,@cimmino); % Train lambda.
options.lambda = lambda;                    % Iterate with
  X1 = cimmino(A,b,1:kmax,[],options);      %  fixed lambda.
options.lambda = 'psi2';                    % Iterate with
  X2 = cimmino(A,b,1:kmax,[],options);      %  psi2 strategy.
options.lambda = 'line';                    % Iterate with
  X3 = cimmino(A,b,1:kmax,[],options);      %  line search.
```

Using AIR Tools – Another Example

```
N = 24;        % Problem size is N-by-N.
eta = 0.05;    % Relative noise level.
kmax = 20;     % Number of of iterations.
```



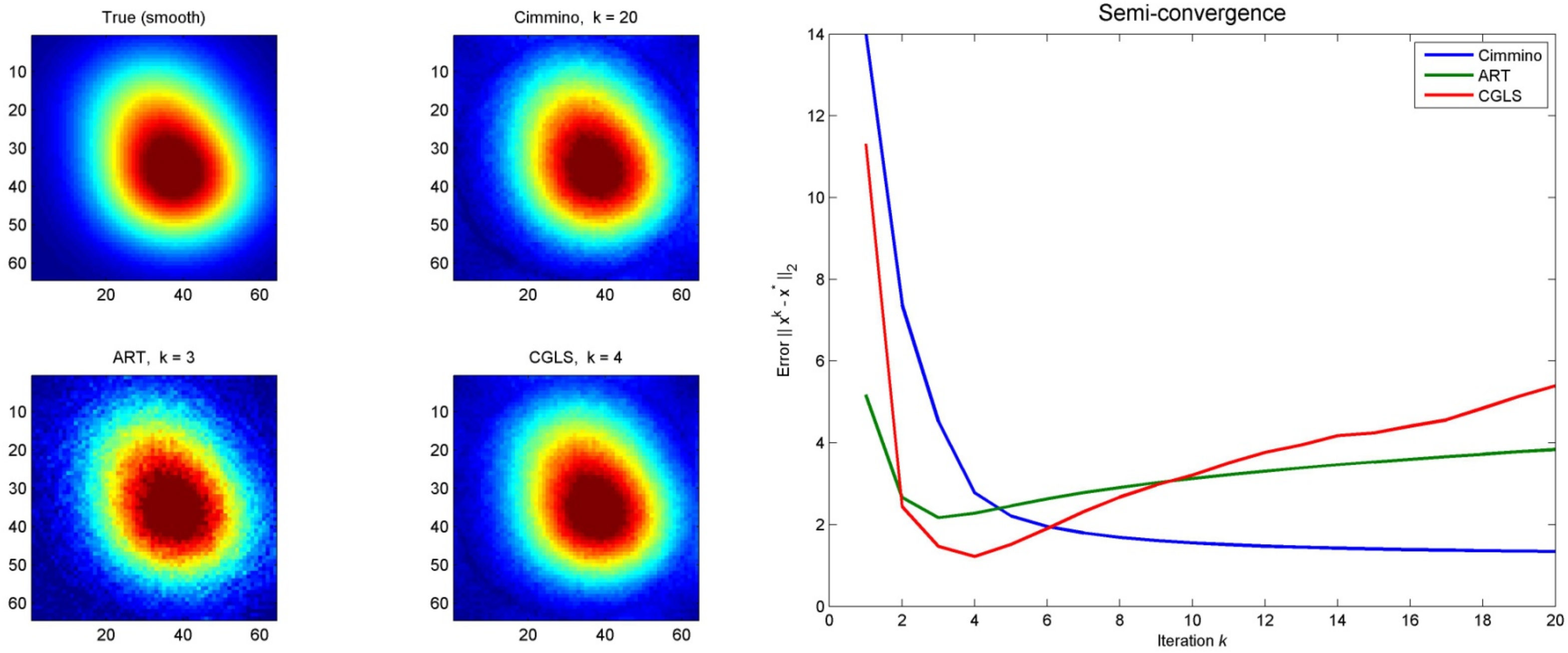

```
[A,bex,xex] = fanbeamtomo(N,10:10:180,32); % Test problem
nx = norm(xex); e = randn(size(bex));       %  with noise.
e = eta*norm(bex)*e/norm(e); b = bex + e;

% Find tau parameter for Discrepancy Principle by training.
delta = norm(e);
options.lambda = 1.5;
tau = trainDPME(A,bex,xex,@randkaczmarz,'DP',delta,2,options);

% Use randomized Kaczmarz with DP stopping criterion.
options.stoprule.type = 'DP';
options.stoprule.taudelta = tau*delta;
[x,info] = randkaczmarz(A,b,kmax,[],options);
k = info(2);   % Number of iterations used.
```

Using AIR Tools – An Third Example

```
N = 64;                        % Problem size.
eta = 0.02;                    % Relative noise level.
k = 20;                        % Number of iterations.
[A,bex,x] = odftomo(N);        % Test problem, smooth image.

% Noisy data.
e = randn(size(bex)); e = eta*norm(bex)*e/norm(e); b = bex + e;

% ART (Kaczmarz) with non-negativity constraints.
options.nonneg = true;
Xart = kaczmarz(A,b,1:k,[],options);

% Cimmino with non-neg. constraints and Psi-2 relax. param. choice.
options.lambda = 'psi2';
Xcimmino = cimmino(A,b,1:k,[],options);

% CGLS followed by non-neg. projection.
Xcgls = cgls(A,b,1:k); Xcgls(Xcgls<0) = 0;
```

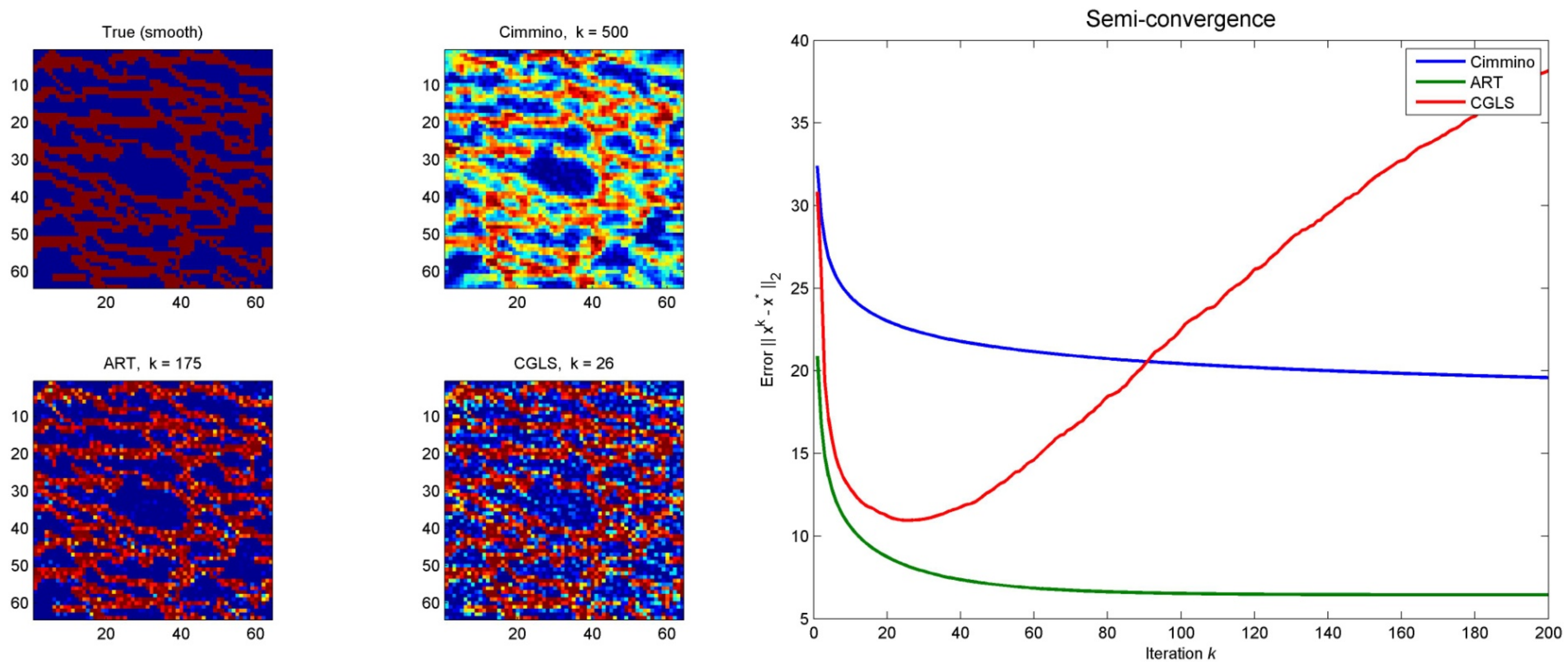
➡ Next page

Results for Smooth Image Example



CGLS gives the best result in just $k = 4$ iterations.

Results for Binary Image Example



ART (Kaczmarz) is the most succesful method here.

Conclusions

- ART and SIRT methods are well suited for tomography.
- Projection on a convex set can easily be incorporated.
- More difficult in incorporate other types of prior information.
- Both methods rely on semi-convergence; it is well understood for the SIRT methods.
- The role of the relaxation parameter is well understood, and we have a strategy that control the noise error.
- We developed a new MATLAB package **AIR Tools** with
  - three methods for choosing the relaxation parameter,
  - three stopping rules, and
  - three test problems.
  - Available from `www.imm.dtu.dk/~pch/AIRtools`.