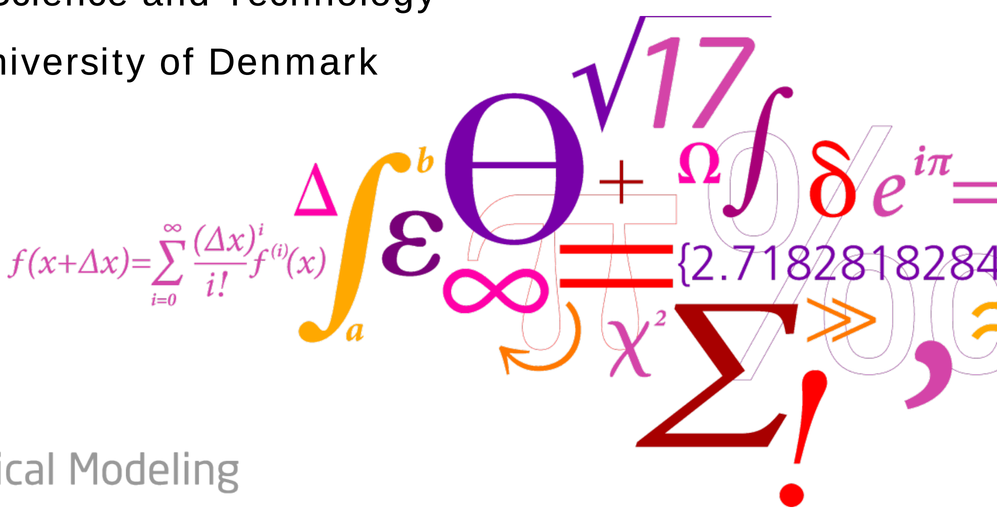
Semi-Convergence and Relaxation Parameters for a Class of SIRT Algorithms

Per Christian Hansen Technical University of Denmark

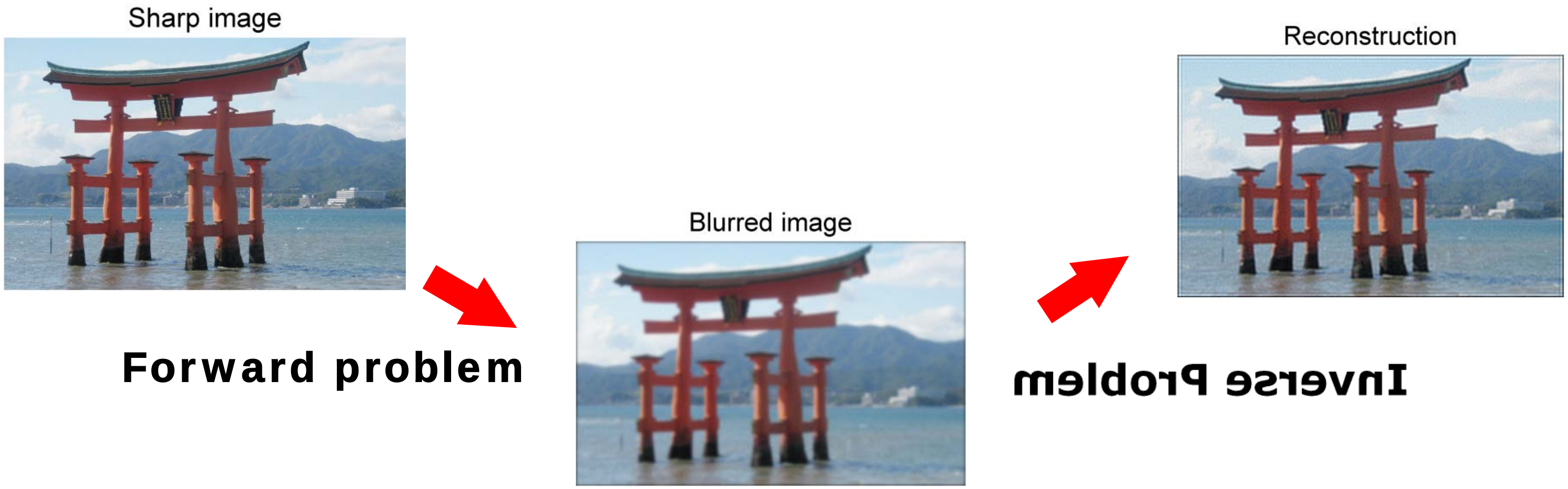
Joint work with:

- Tommy Elfving Linköping University, Sweden
- Touraj Nikazad Iran University of Science and Technology
- Maria Saxild-Hansen Technical University of Denmark



DTU Informatics
Department of Informatics and Mathematical Modeling

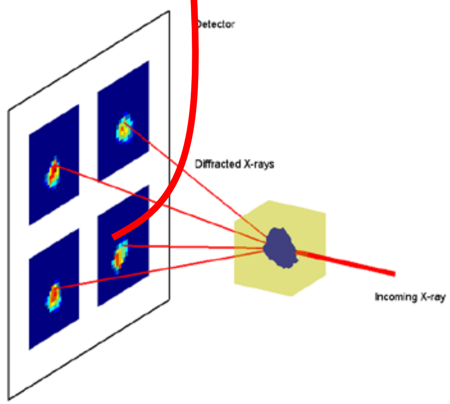
Overview of Talk

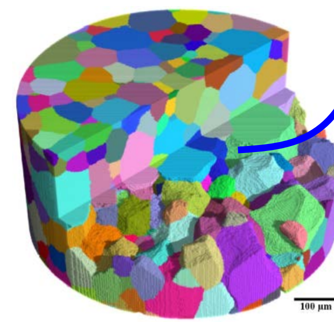


- Inverse problems and reconstruction algorithms
- Iterative SIRT methods and their semi-convergence
- Strategies for the relaxation parameter (step size)
- A few results
- If time permits: AIR Tools – a new MATLAB® package

Inverse Problems

Goal: use measured **data** to ***compute*** "hidden" **information**.





Blurring process

Sharp image

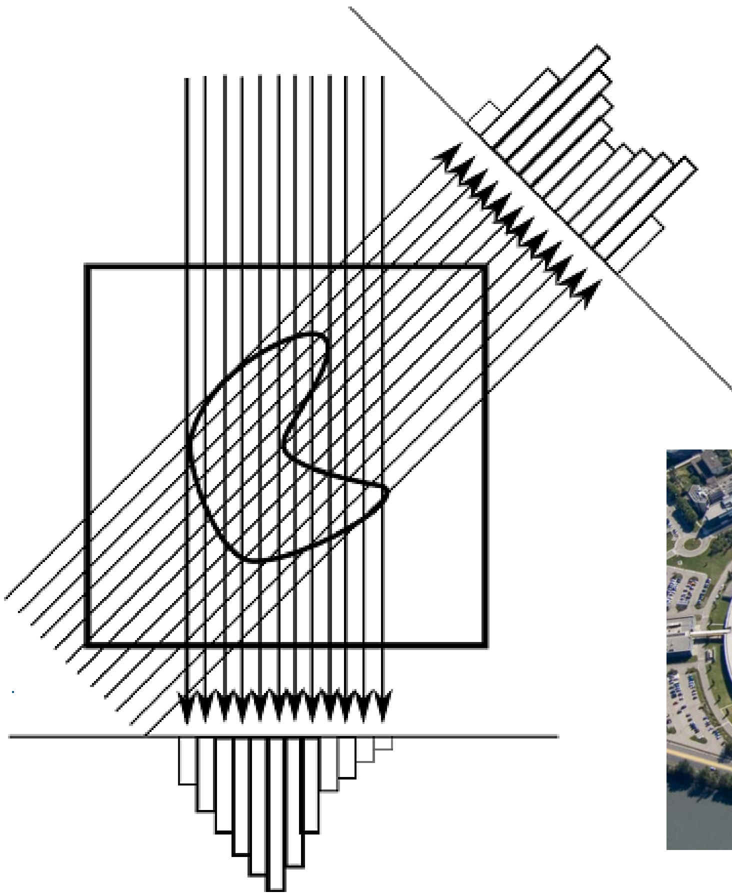
Our model: $A\,x = b$

Data / blurred image

Tomography = Our Main Application Area

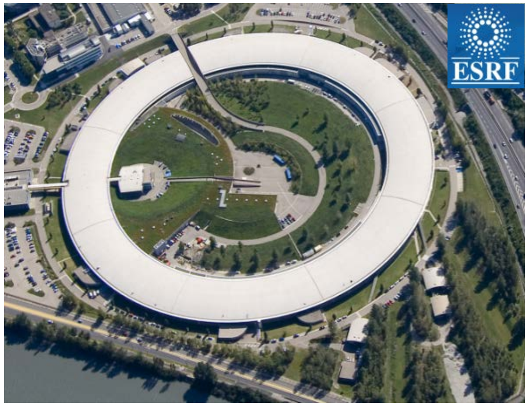
Image reconstruction

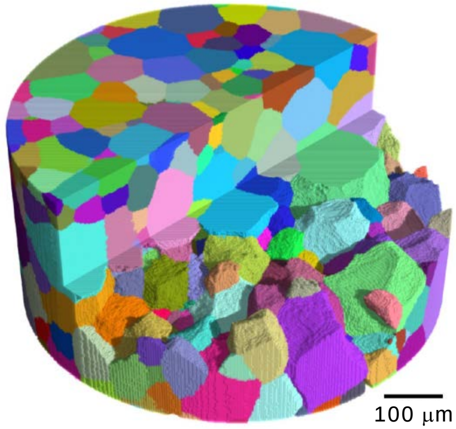
from projections



Medical scanning

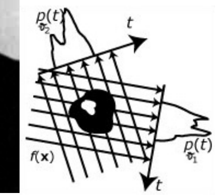
Mapping of metal grains





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

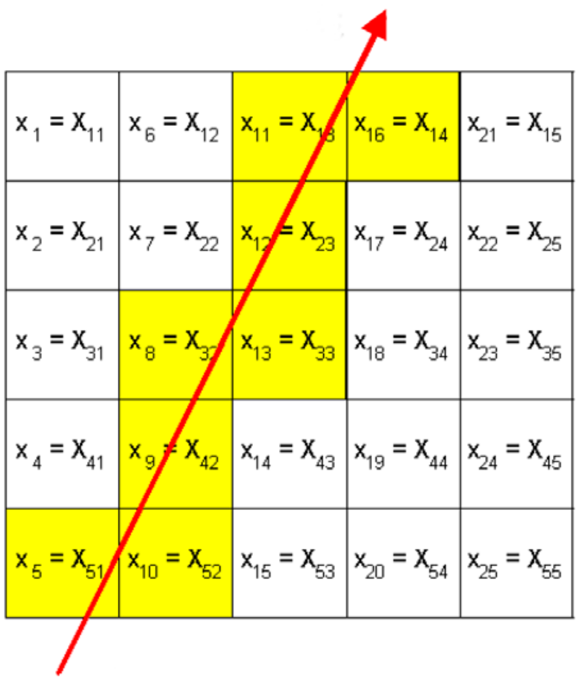
Allan M Cormack and Godfrey Newbold Hounsfield

for the "development of computer assisted tomography".

Setting Up the Algebraic Model

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

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



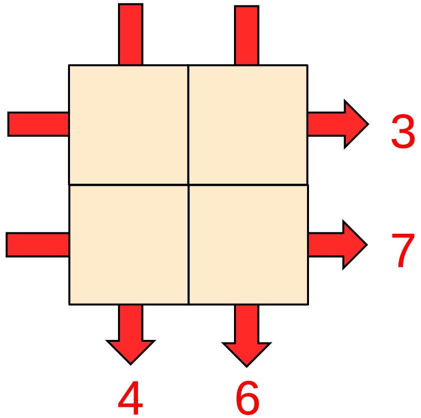
Discretization leads to a large, sparse, ill-conditioned system:

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

Geometry (A), Image (x), Projections (b), Noise (e)

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

Analogy: the “Sudoku” Problem – 数独



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

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

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

Prior: solution is integer and non-negative

0	3
4	3

1	2
3	4

2	1
2	5

3	0
1	6

Some Large-Scale Inversion Algorithms

Transform-Based Methods

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

Example: the inverse Radon transform for tomography.

Krylov Subspace Methods

Use the forward model to produce a Krylov subspace →
inversion amounts to projecting on this "signal subspace" & using prior information. Examples: CGLS, RRGMRES.

Algebraic Iterative Methods

Formulate the forward problem as a discretized problem →
inversion amounts to solving $A\,x = b$ using algebraic properties of A & using prior information.

This work

Some Algebraic Iterative 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.

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).

Making the methods useful

- Relaxation parameter (step length) choice.
- Stopping rules.
- Nonnegativity constraints.

This work

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 $M = \mathrm{diag}(\text{column sums})^{-1}$.

Semi-Convergence of the SIRT Methods

During the first iterations, the iterates x^k capture the "important" information in the noisy right-hand side b.

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

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

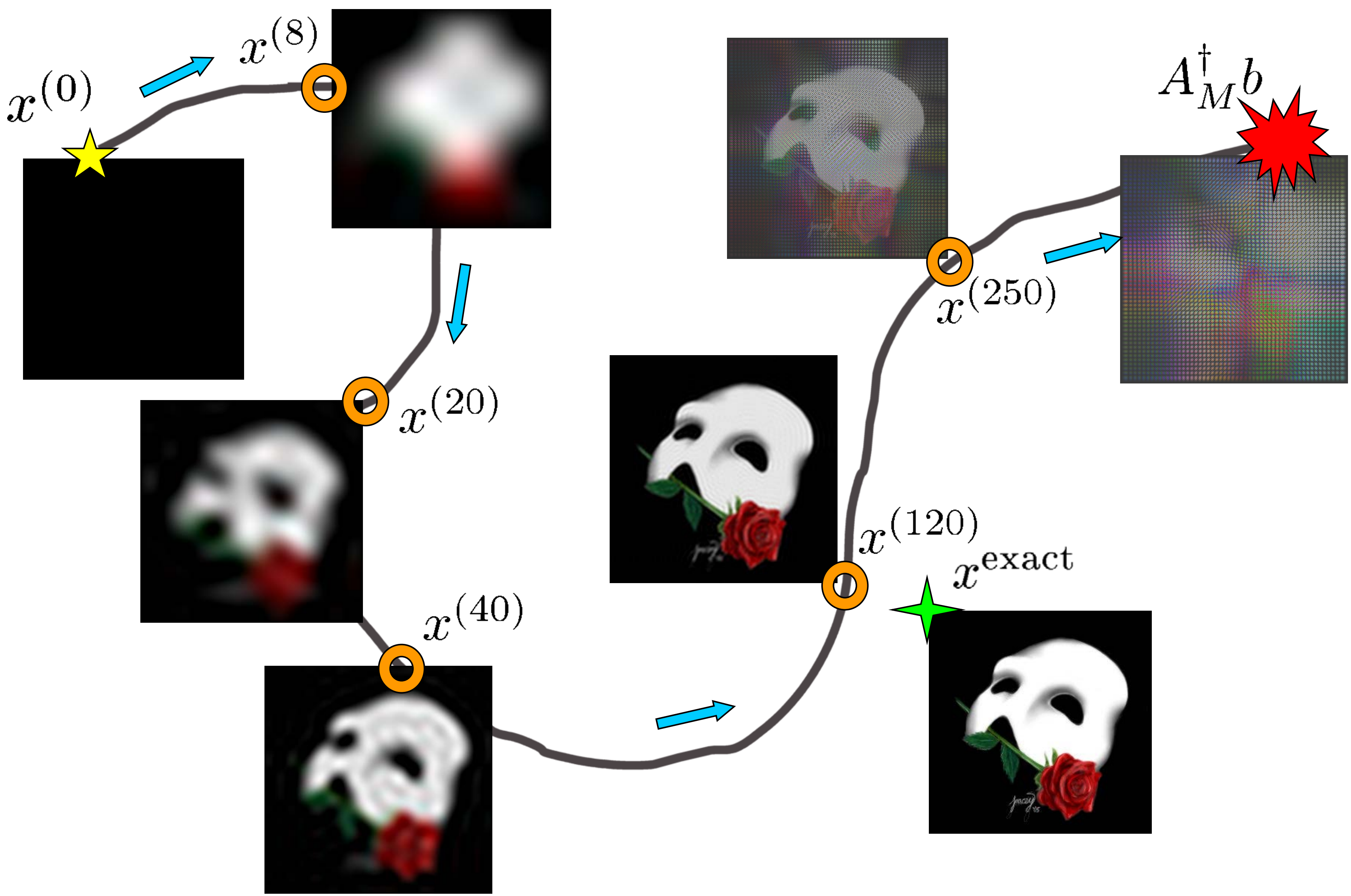
- Now the iterates x^k diverge from the exact solution and they approach the undesired solution $A^{-1}b$ or $A^{\dagger}b$.

The iteration number k plays the role of the regularization parameter.

This behavior is called *semi-convergence*.

- F. Natterer, *The Mathematics of Computerized Tomography* (1986)
- 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. 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

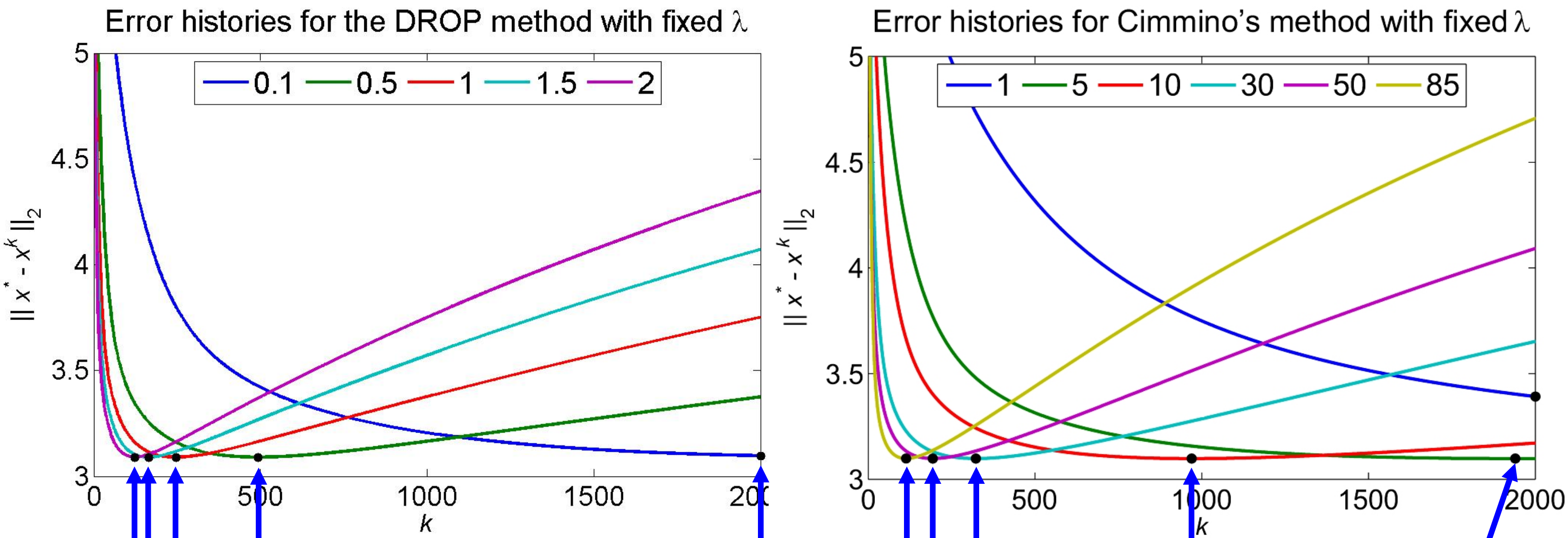


Another Look at Semi-Convergence

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

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

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



The <u>minimum error</u> is *independent* of both λ and the method.

Analysis of Semi-Convergence

Let $\bar{x}$ be the solution to the noise-free problem:

$$\bar{x} = \text{argmin}_{x \in \mathcal{C}} {}^1/_2 \|A\,x - \bar{b}\|_M^2, \qquad \bar{b} = \text{pure data}$$

and let $\bar{x}^k$ denote the iterates when applying SIRT to $\bar{b}$. Then

$$\|x^k - x^*\|_2 \;\square\; \|x^k - \bar{x}^k\|_2 + \|\bar{x}^k - \bar{x}\|_2 \;.$$

Noise error

Iteration error

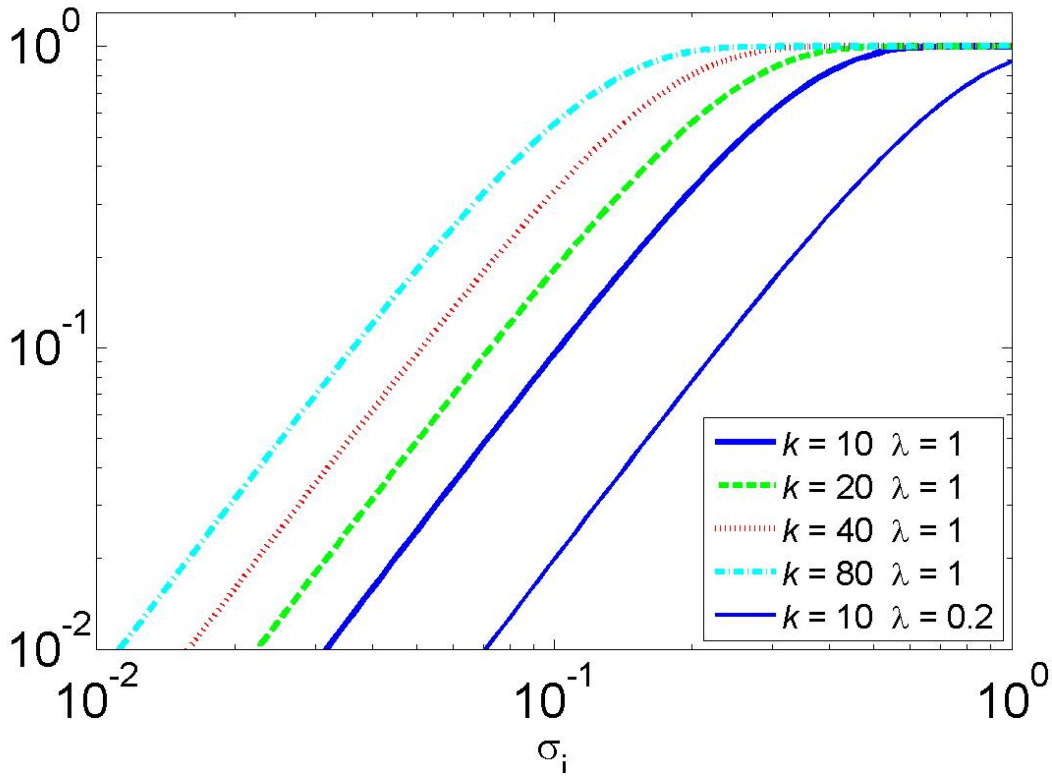
We need the SVD: $\boxed{M^{1/2} A = U\,\Sigma\,V^T}$ Assume $\text{rank}(A) = n$.

The unprojected case is "easy;" x^k is a filtered SVD solution:

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

The Behavior of the Filter Factors

Filter factors $\varphi_i^{[k]} = 1 - \left(1 - \lambda\,\sigma_i^2\right)^k$



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

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

Projected Alg. Noise Error (proof: see paper)

The noise error in projected SIRT is bounded above by

$$\|x^k - \bar{x}^k\|_2 \le \frac{\sigma_1 \lambda_0}{\sigma_n \lambda_{k-1}} \, \Psi^k(\lambda_{k-1}) \, \|M^{1/2} \delta b\|_2,$$

with

$$\Psi^k(\lambda) \equiv \frac{1 - (1 - \lambda \sigma_n^2)^k}{\sigma_n}.$$

When $\lambda_k = \lambda$ for all k we obtain

$$\|x^k - \bar{x}^k\|_2 \le \frac{\sigma_1}{\sigma_n} \, \Psi^k(\lambda) \, \|M^{1/2} \delta b\|_2,$$

and as long as $\lambda \sigma_n^2 \ll 1$ we have

$$\|x^k - \bar{x}^k\| \approx \lambda \, k \, \sigma_1 \|M^{1/2} \delta b\|_2,$$

showing that k and λ play the same role for suppressing the noise.

Projected Alg. Iteration Error (proof: see paper)

The iteration error in projected SIRT is bounded above by

$$\|\bar{x}^k - \bar{x}\|_2 \leq \sigma_n \Phi^k(\lambda_{k-1}) \, \|x^0 - \bar{x}\|_2,$$

with

$$\Phi^k(\lambda) \equiv \frac{(1-\lambda\sigma_n^2)^k}{\sigma_n}.$$

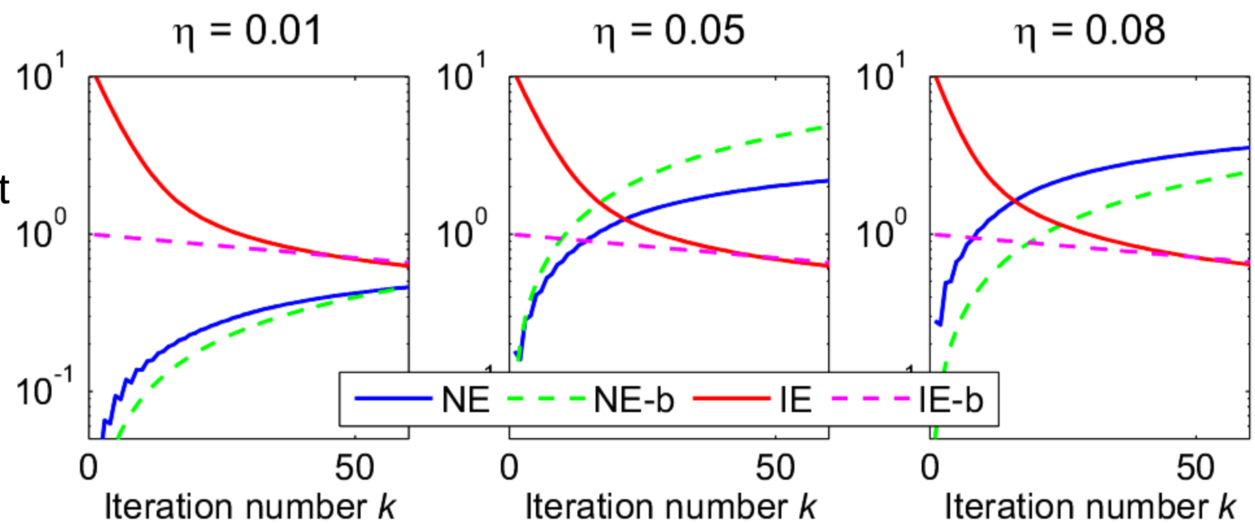
Our bound have pessimistic factors, but **track** well the actual errors:

NE: actual noise error

NE-b: our bound without the factor σ_1/σ_n

IE: actual iteration error

IE-b: our bound without the factor $\|x^0 - \bar{x}\|_2$

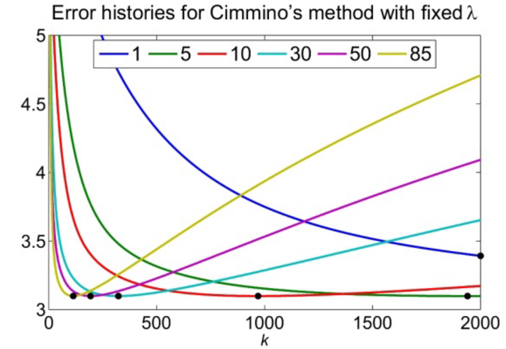


Choosing the SIRT Relaxation Parameter

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

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.



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:

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

Preparation for More Insight …

The function (which appears in the analysis)

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

has a unique real root $\zeta_k \in (0,1)$. The roots satisfy

$$0 < \zeta_k < \zeta_{k+1} < 1 \qquad \text{and} \qquad \lim_{k\to\infty} = 1$$

k	ζ_k	k	ζ_k	k	ζ_k	k	ζ_k	k	ζ_k
2	0.3333	7	0.8156	12	0.8936	17	0.9252	22	0.9424
3	0.5583	8	0.8392	13	0.9019	18	0.9294	23	0.9449
4	0.6719	9	0.8574	14	0.9090	19	0.9332	24	0.9472
5	0.7394	10	0.8719	15	0.9151	20	0.9366	25	0.9493
6	0.7840	11	0.8837	16	0.9205	21	0.9396	26	0.9513

Parameter-Choice: Limit the Noise Error

Assume that $0 < \lambda_{i-1} \square \lambda_i$ in steps $1, \ldots, k-1$; then

$$\|x^k - \bar{x}^k\|_2 \square \frac{\sigma_1 \lambda_0}{\sigma_n \sqrt{\lambda_{k-1}}} \frac{1-\zeta_k^k}{\sqrt{1-\zeta_k}} \|M^{1/2} \delta b\|_2$$

<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

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

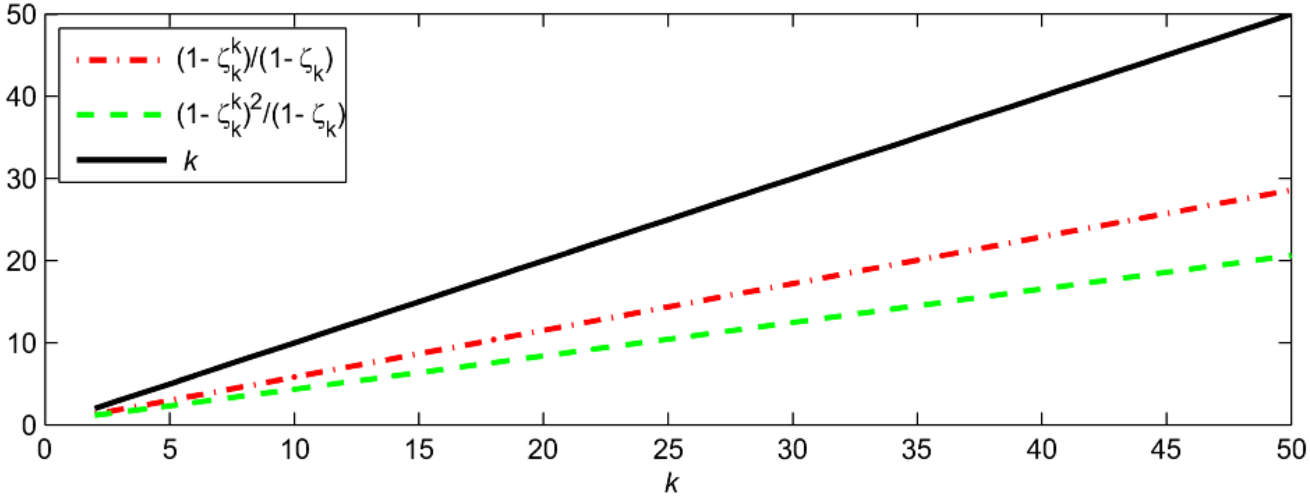
Both are diminishing: $\lambda_k \to 0$ such that $\sum_k \lambda_k = \infty$.

Our New Strategies: What we Achieve

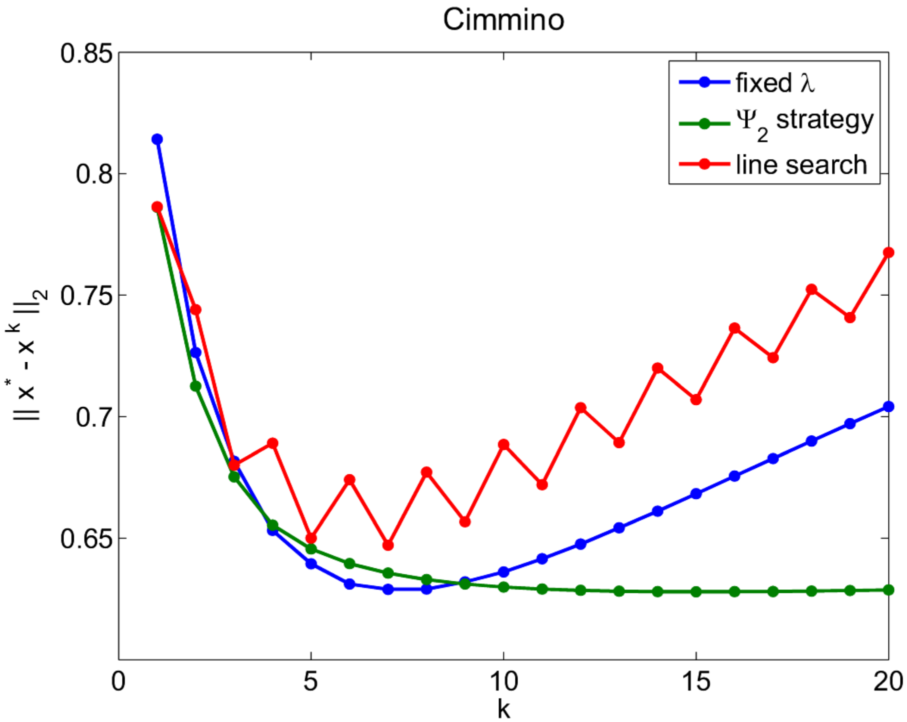
As a result:

$$\|x^k - \bar{x}_k\|_2 \quad \square \quad \frac{\sigma_1^2 \lambda_0}{\sigma_n \sqrt{2}} \, \frac{1-\zeta_k^k}{1-\zeta_k} \, \|M^{1/2} \delta b\|_2 \quad \text{for strategy } \Psi_1$$

$$\|x^k - \bar{x}_k\|_2 \quad \square \quad \frac{\sigma_1^2 \lambda_0}{\sigma_n \sqrt{2}} \, \frac{(1-\zeta_k^k)^2}{1-\zeta_k} \, \|M^{1/2} \delta b\|_2 \quad \text{for strategy } \Psi_2$$



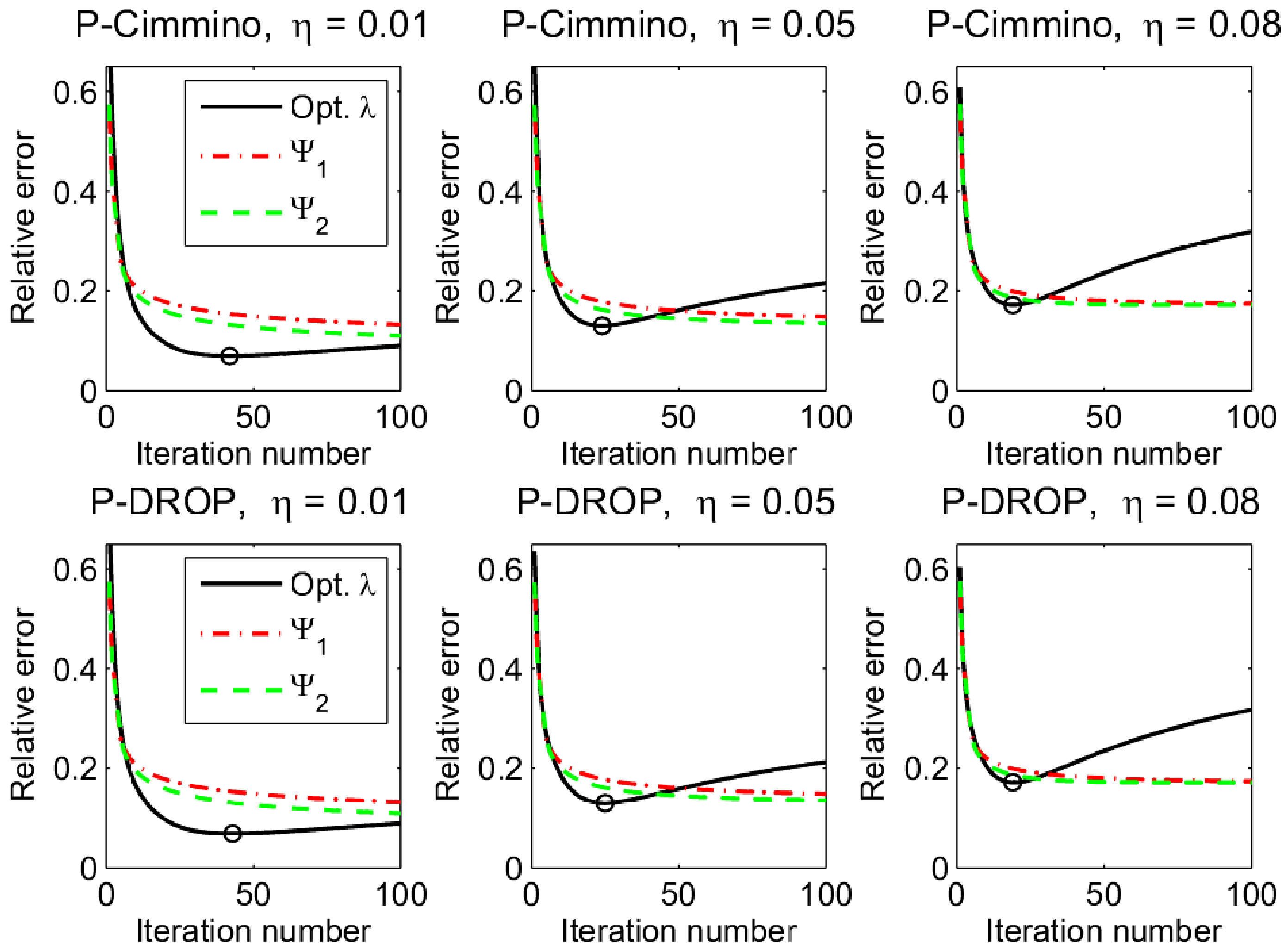
Error Histories for Cimmino Example



All three strategies give fast semi-convergence:

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

Numerical Results (SNARK model problem)



Conclusions

- We have verified the observed simiconvergence of the standard and the projected SIRT methods.
- We proposed two new strategies for choosing λ_k.
- Our strategies control the noise component of the error.
- In case of noise-free data our strategies give convergence to the problem $\min_{x \in \mathcal{C}} \|A\, x - b\|_M^2$.
- Our strategies also work for consistent and inconsistent systems, for rank-deficient matrices, and SIRT methods with $T \neq I$.
- They are implemented in the MATLAB package AIR Tools.