Numerical Methods for Ill-Posed Problems

Misha Kilmer

Tufts University, Medford, MA

Outline

- Definition of DIPP's, sample problem
- On the need for regularization
- Krylov subspace regularization
- Choosing regularization parameters
- Preconditioning, structured matrices
- Hybrid approaches
- Summary

Background

An problem is ill-posed if it is not unique or it is not a continuous function of the data [Hadamard, '23].

First-kind Fredholm integral equations

$$\int_\Omega K(s,t)f(t)dt = g(s),$$

are notoriously ill-posed.

Motivation

Why linear ill-posed problems?

- Applications where model is appropriate:
  - Image deblurring
  - Computerized tomography
- Tractable
- Initial guesses for nonlinear inverse problems
- Nonlinear problems may need specialized regularization techniques

Discrete Ill-Posed Problem

Solve for f_{true}, given A, g and the model

$$Af_{true} = g_{true} + e = g,$$

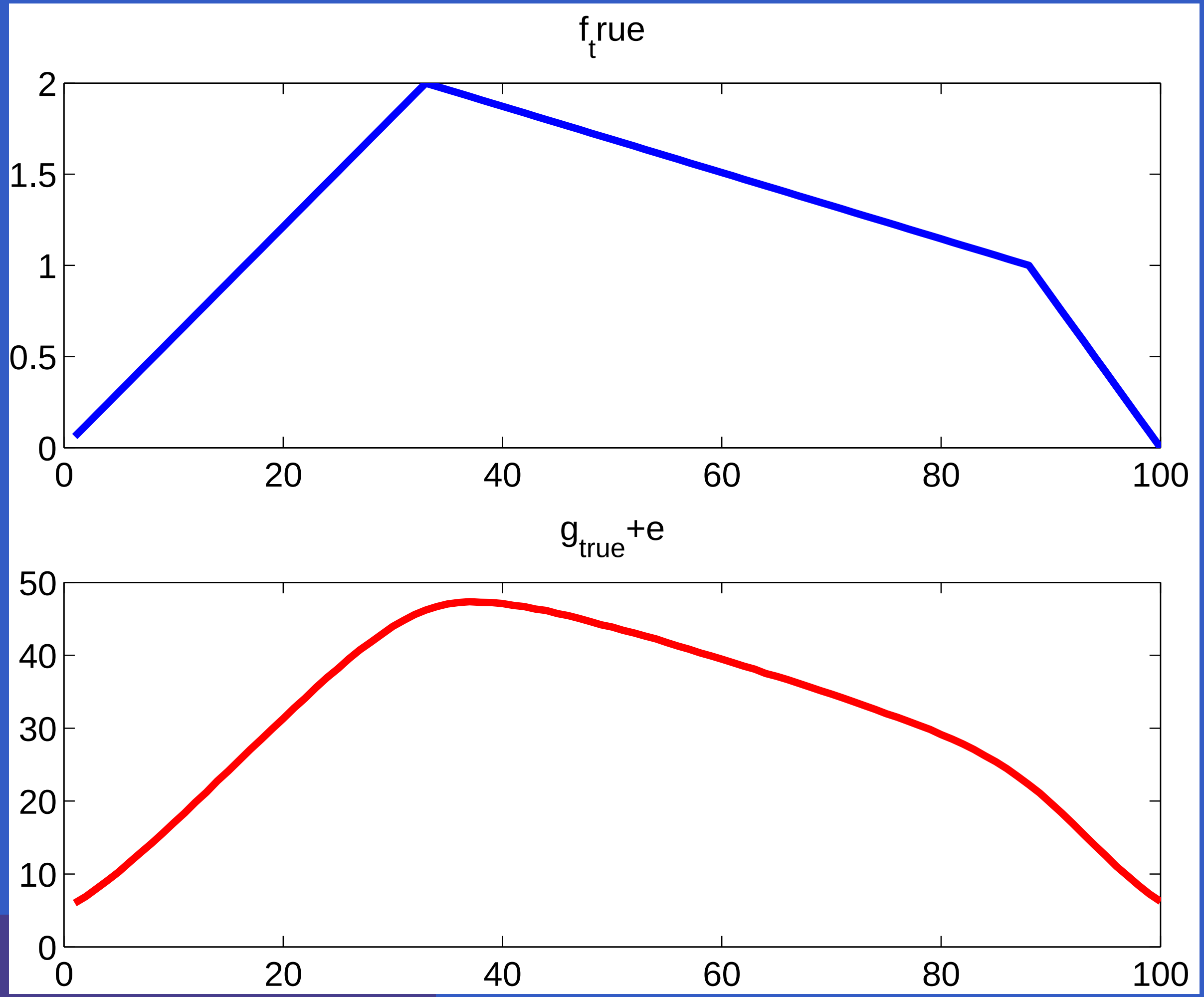
where $A \in \mathbb{R}^{m \times n}, m \geq n$ is full rank.

Properties:

- Decaying singular values, no gap
- Noise is unknown (white), but $\|e\|/\|g_{true}\| < 1$
- m, n are large
- Discrete Picard condition holds

Gravity Test Problem

$f(t)$: mass distribution depth d,
$g(s)$: vertical component of gravity field



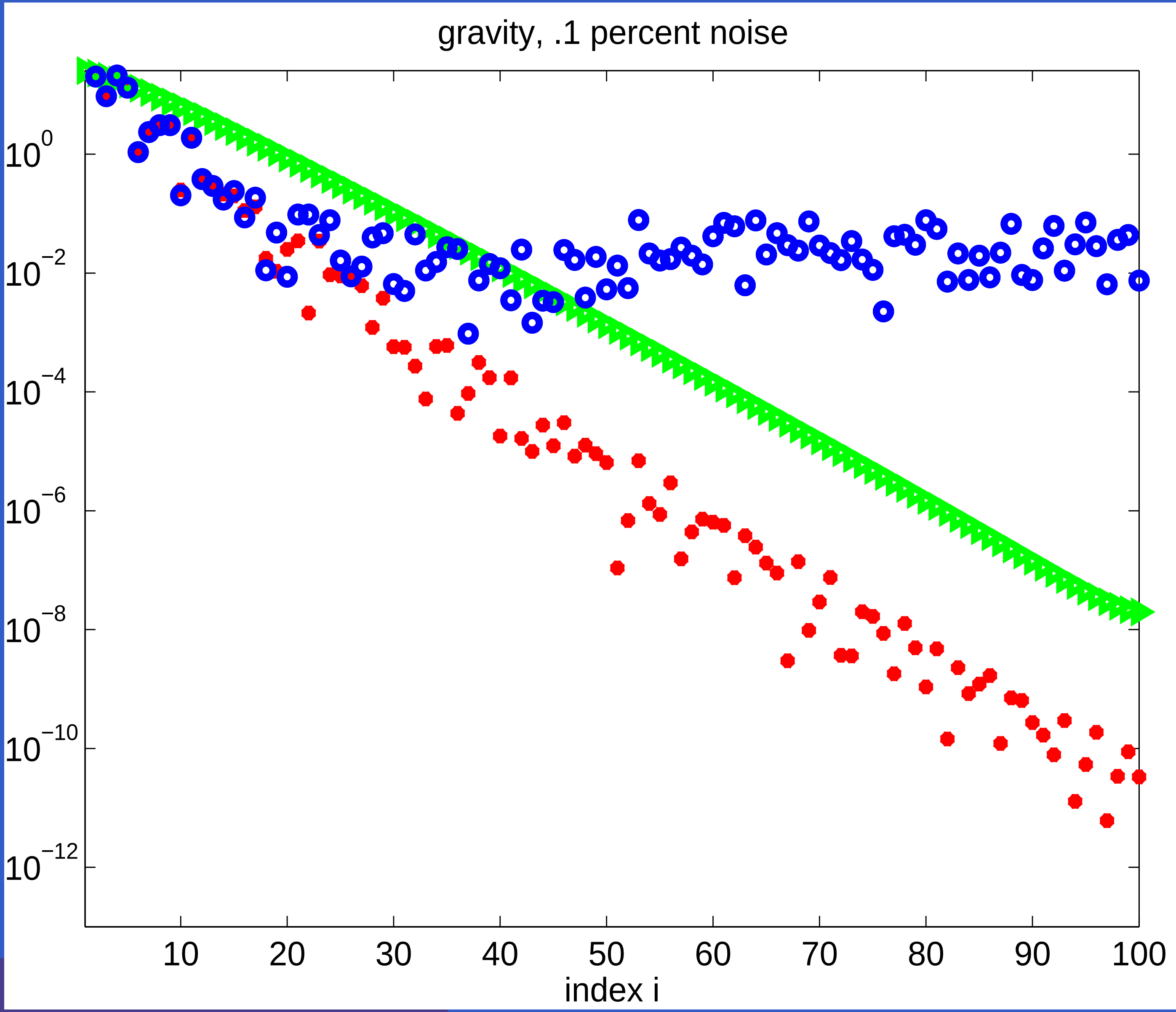
Need for Regularization

Let $A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$.

The exact solution to $Af = g = g^{true} + e$:

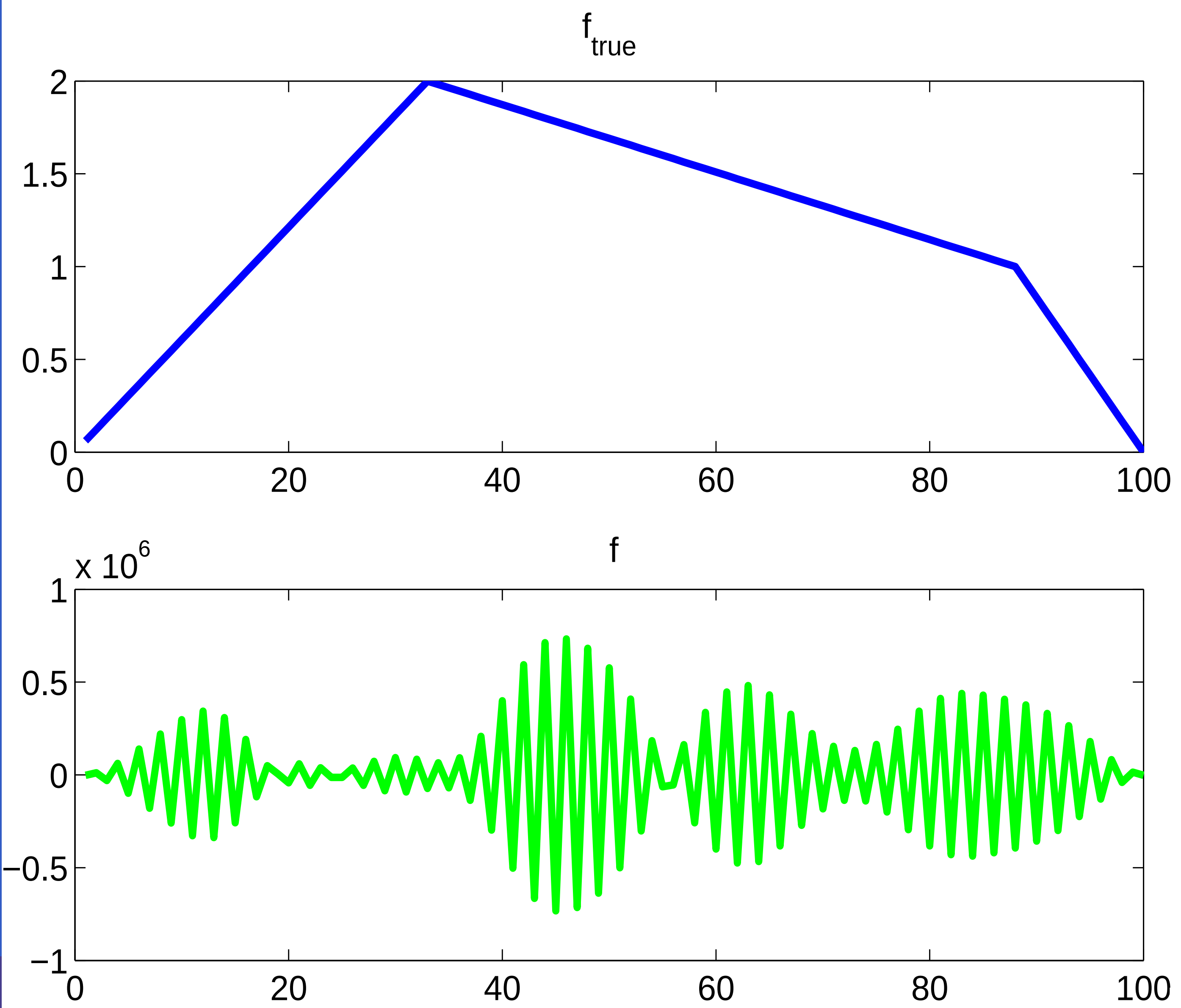
$$f = \sum_{i=1}^{n} \frac{u_i^T g^{true}}{\sigma_i} v_i + \frac{\overbrace{u_i^T e}^{\approx c}}{\sigma_i} v_i$$

Picard Plot



gravity, .1 percent noise
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10
20
30
40
50
60
70
80
90
100
index i

Exact Solution



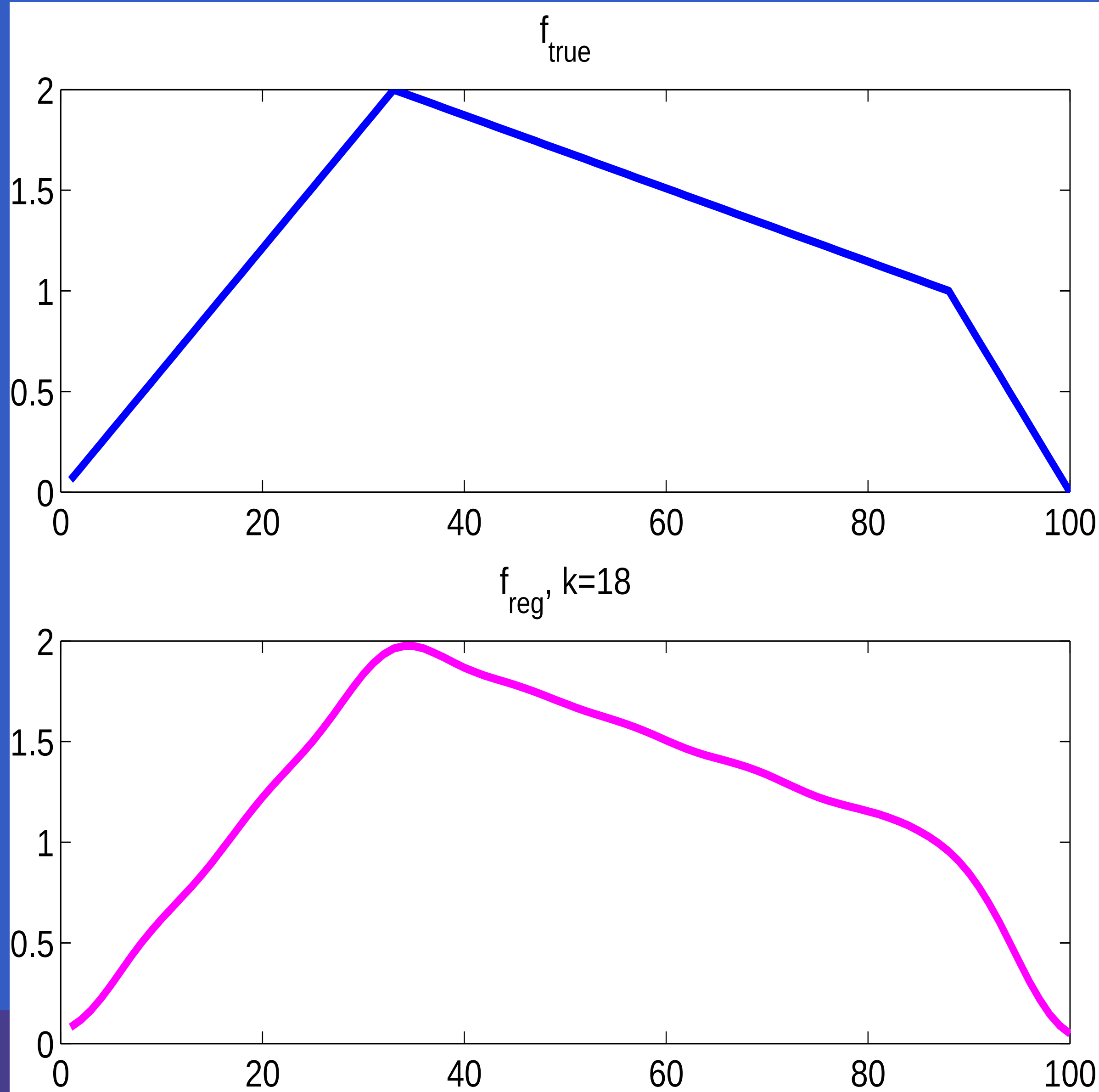
Truncated SVD

Instead, take the regularized solution

$$f_{reg} = \sum_{i=1}^{k} \frac{u_i^T g}{\sigma_i} v_i$$

SVD unrealistic for large problems!

Optimal TSVD Solution



Krylov Subspace Methods

$$K_k(B,v) = \text{span}\{v, Bv, B^2v, \ldots, B^{k-1}v\}$$

- Conjugate Gradient (SPD A):

$$f^{(k)} = \arg \min_{z \in \mathcal{K}_k(A,g)} \|z - f\|_A$$

- MINRES (A symmetric)

$$f^{(k)} = \arg \min_{z \in \mathcal{K}_k(A,g)} \|Az - g\|_2$$

- LSQR (or CGLS)

$$f^{(k)} = \arg \min_{z \in \mathcal{K}_k(A^TA,A^Tg)} \|Az - g\|_2$$

Krylov Subspace Methods

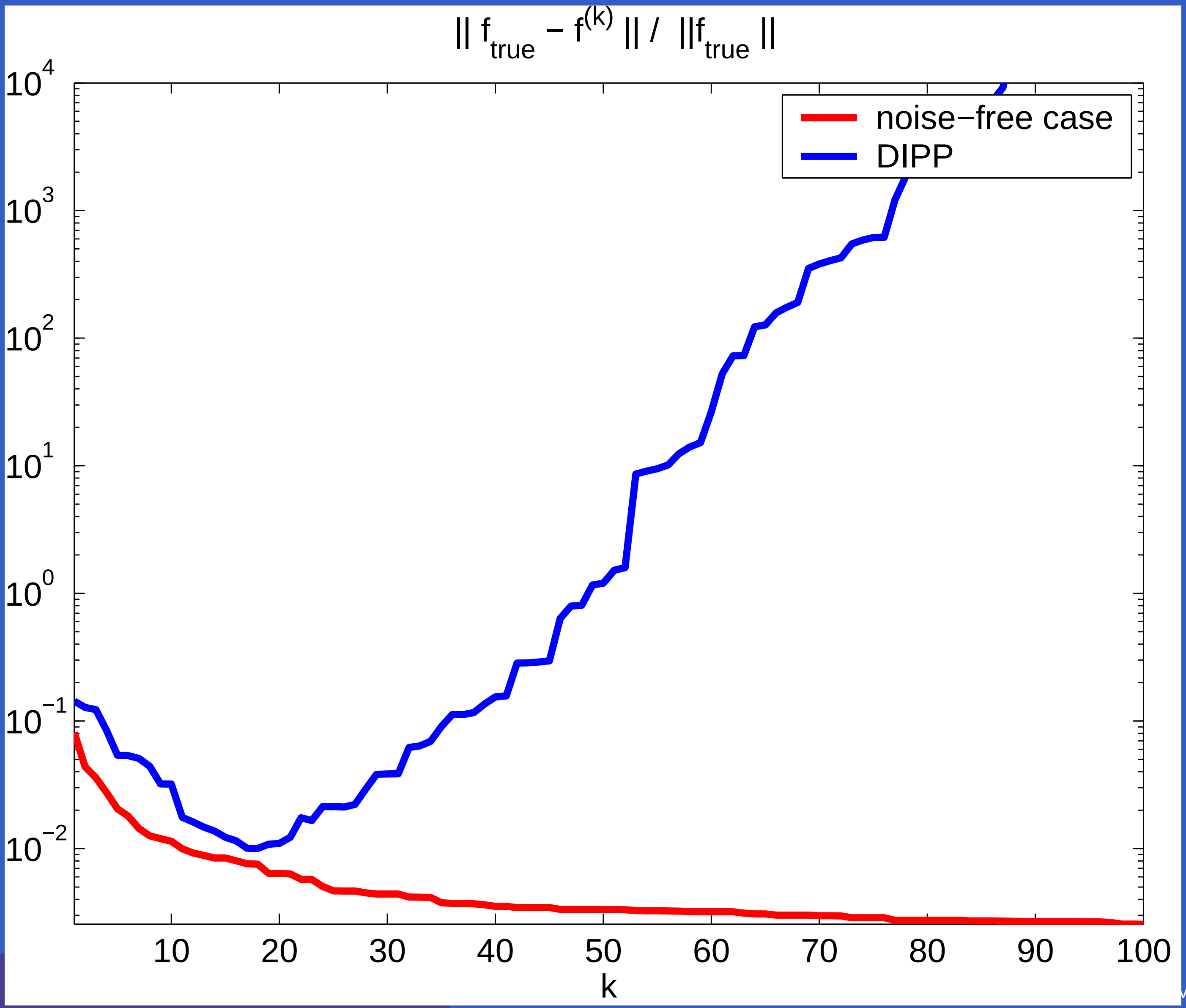
- Each iteration “costs" 1 or 2 mat-vecs and a few dot-products, saxpys.
- Iterates, basis vectors updated via short-term recurrences $\Rightarrow$ low storage.
- Convergence rate (to f) depends on clustering of spectrum.

LSQR as a regularization method

LSQR minimizes the norm of the residual, $Af^{(k)} - g$, at every iteration $\Rightarrow$ decreasing function of k.

Relative errors $\dfrac{\|f_{true} - f^{(k)}\|}{\|f_{true}\|}$ are another story!

Rel. Errors, noisy & noise-free cases



LSQR

$$f^{(k)} = \sum_{i=1}^{n} \underbrace{\phi^{(k)}(\sigma_i^2)}_{\text{filter factors}} \frac{u_i^T g}{\sigma_i} v_i$$

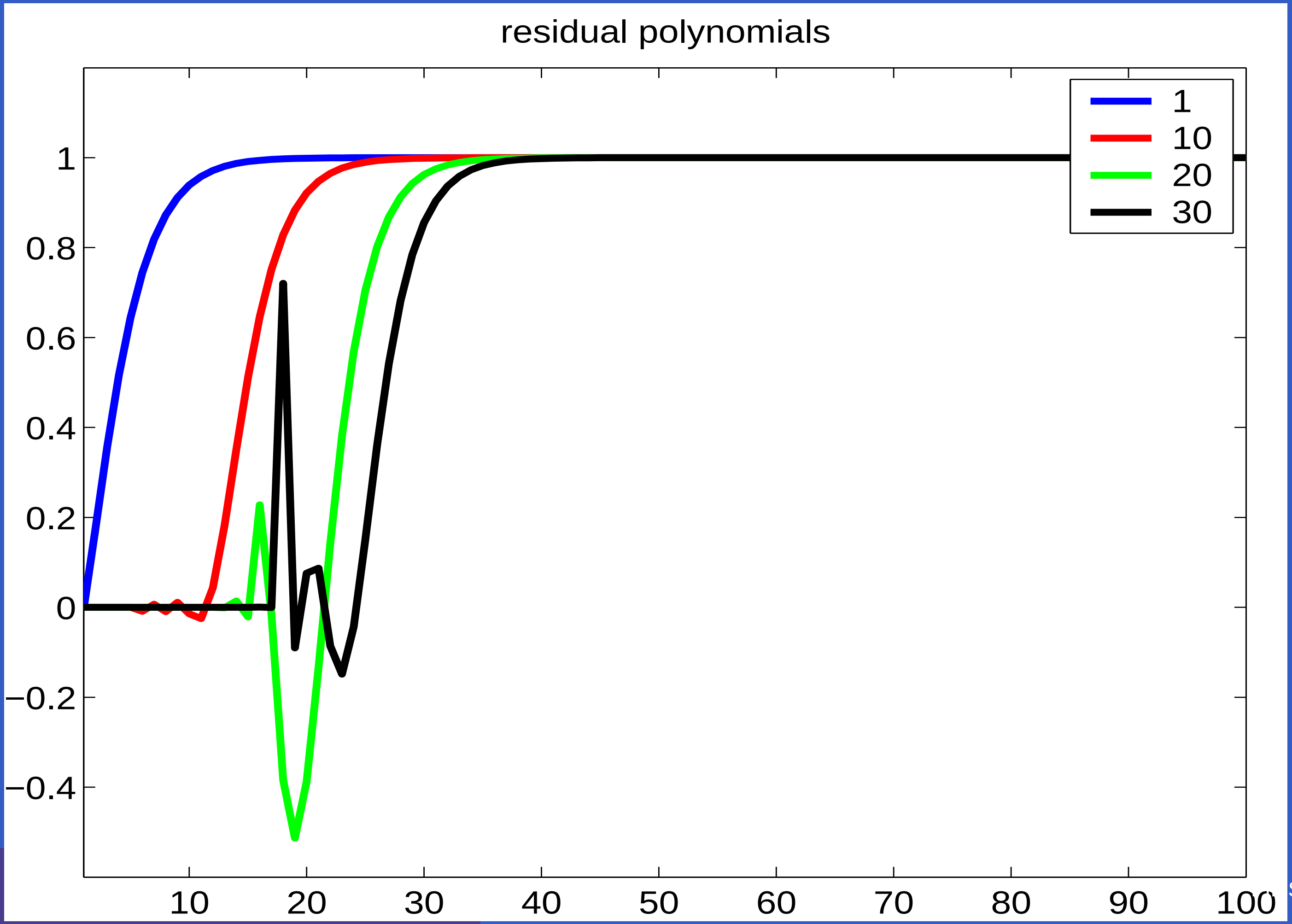
where $\phi^{(k)} \in \Pi^k$.

- Study of convergence in regularization case different from usual analysis.
- See [Hanke '95] analysis in Hilbert space setting.
- Use analysis of the residual [K. and Stewart '99, K. '00].

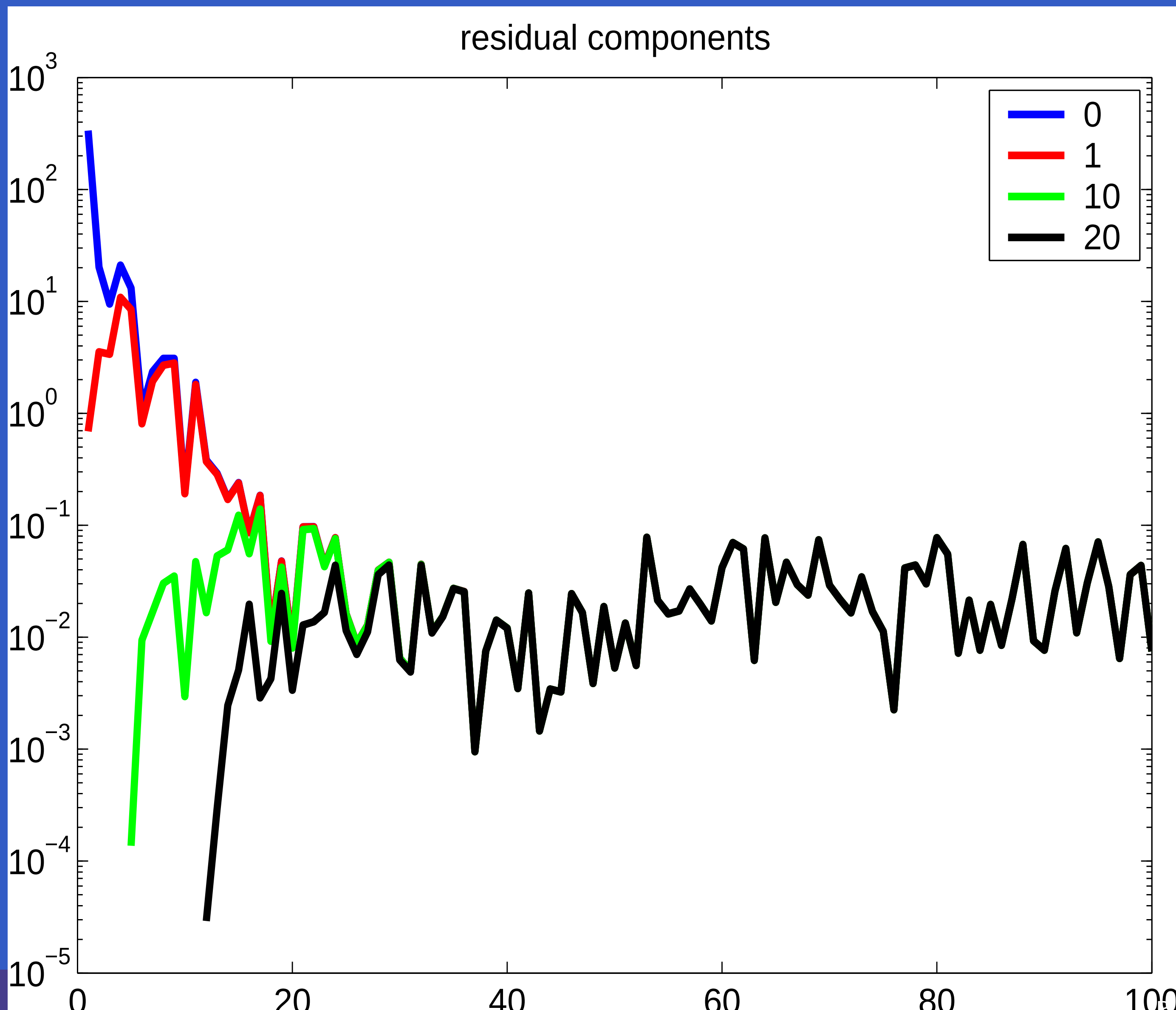
Residual Polynomials

Residual poly $p^{(k)}(t) = 1 - (\phi^{(k)}(t))$, & $p^{(k)}(0) = 1$,

$$\|r^{(k)}\| = \|p^{(k)}(AA^*)g\| = \|p^{(k)}(\Sigma\Sigma^T)U^T g\|$$



Plots of $|U^T r^{(k)}|$



Regularization and LSQR

Summary of regularizing properties:

- Residual polynomial **must reduce** residual norm at each step.
- There is more to reduce over the signal subspace early on.
- A root near a small singular value would cause the residual norm to increase.
- Once residual norm falls much below $\|e\|$, solution becomes contaminated.

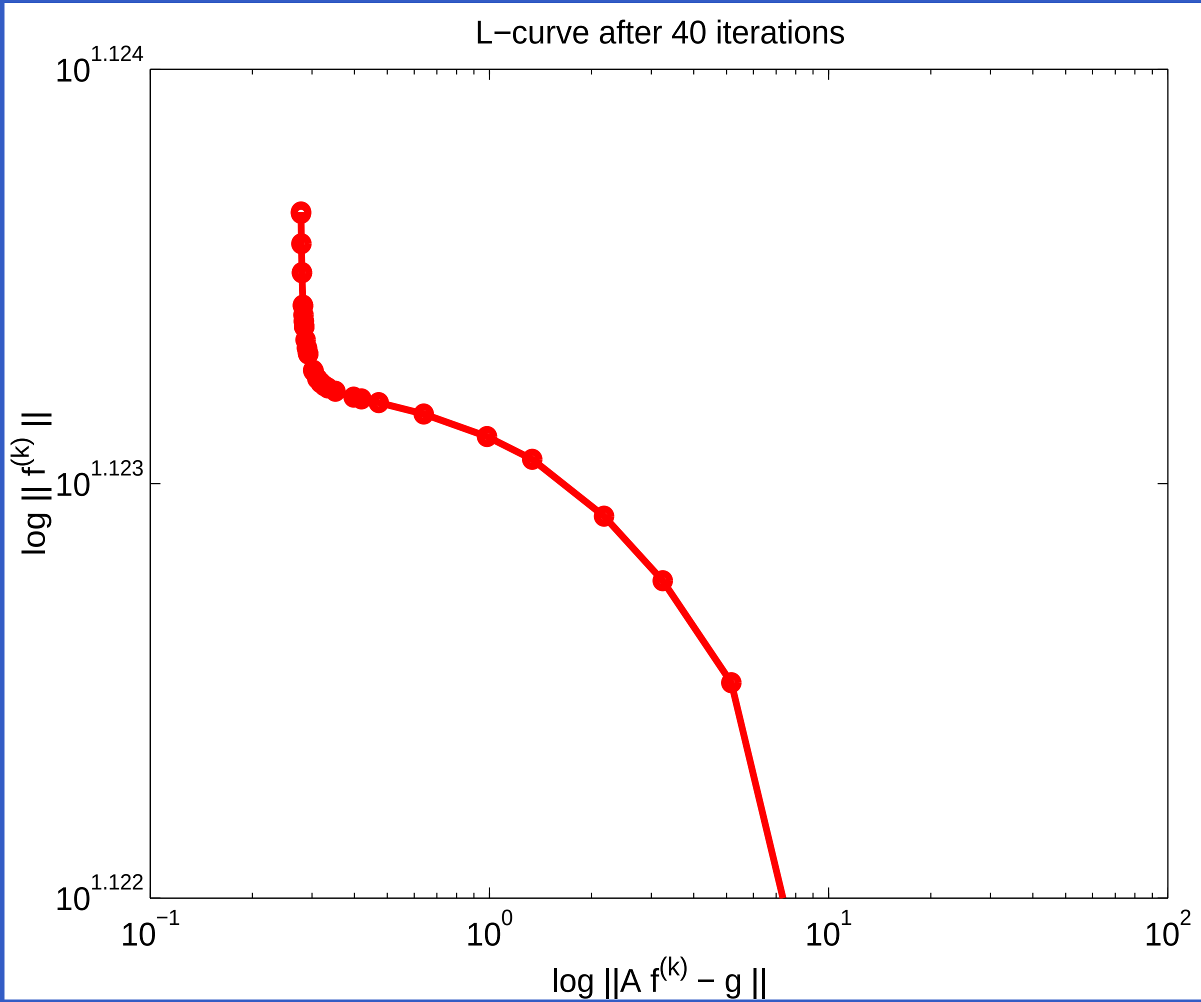
Regularization Parameter k

Incomplete list of options:

- Discrepancy [Morozov '66]:
  Choose k so $\|Af^{(k)} - g\| \approx \|e\|$
- GCV [Golub, Heath, Wahba, '79]:
  minimize $\|Af^{(k)} - g\| / \mathcal{T}(k)$
- L-curve [Hansen & O'Leary '93]:
  "Corner" of $(\log \|Af^{(k)} - g\|, \log \|f^{(k)}\|)$
- CSD [Hansen, K. & Kjeldsen, '02]:
  $\min_k \max($csd$(f^{(k)}, Af^{(k)} - g))$.

Gravity, L-curve

corner at $k = 18$ is optimal



Preconditioning

Assume, for simplicity, A is square:

- $M^{-1}Af = M^{-1}g$
- $AM^{-1}f = g, \qquad y = Mx$
- $M_1^{-1}AM_2^{-1}f = M_1^{-1}g, \qquad y = M_2x$

Inverses usually formed implicitly:

$$M^{-1}v = y, \qquad \Longrightarrow \qquad v = My$$

so “solves" with M must be fast, low storage.

Preconditioning

- In non-noisy case:
  - Preconditioned matrix has singular values clustered away from 0.

    e.g. if $M \approx A,$ then $M^{-1}A \approx I$

- In noisy case:
  - Only cluster part of spectrum corresponding to signal subspace!
  - Preconditioning should not mix signal and noise subspace!

Preconditioning for Regularization

Consider

$$M^{-1}Af = M^{-1}g.$$

If $M \approx A$, then $M^{-1}g \approx A^{-1}g$, which is contaminated by noise!

An “Optimal" Preconditioner

Let $A = U\Sigma V^T = U \begin{bmatrix} \Sigma_j & 0 \\ 0 & \Sigma_{n-j} \end{bmatrix} V^T,$

$$M = U \begin{bmatrix} \Sigma_j & 0 \\ 0 & I \end{bmatrix} V^T.$$

Then

$$M^{-1}A = V \begin{bmatrix} I & 0 \\ 0 & \Sigma_{n-j} \end{bmatrix} V^T$$

.

Preconditioning and Matrix Structure

Toeplitz: $$\begin{bmatrix} t_0 & t_1 & \dots & \dots & t_{j-1} \\ t_{-1} & t_0 & t_1 & \dots & t_{j-2} \\ \vdots & \ddots & t_0 & \ddots & \ddots \\ t_{1-j} & \ddots & \ddots & \ddots & t_0 \end{bmatrix}.$$

BTTB: $$\begin{bmatrix} T_0 & T_1 & \dots & \dots & T_{j-1} \\ T_{-1} & T_0 & T_1 & \dots & T_{j-2} \\ \vdots & \ddots & T_0 & \ddots & \ddots \\ T_{1-j} & \ddots & \ddots & \ddots & T_0 \end{bmatrix}, T_l \text{ is Toeplitz.}$$

Preconditioning and Matrix Structure

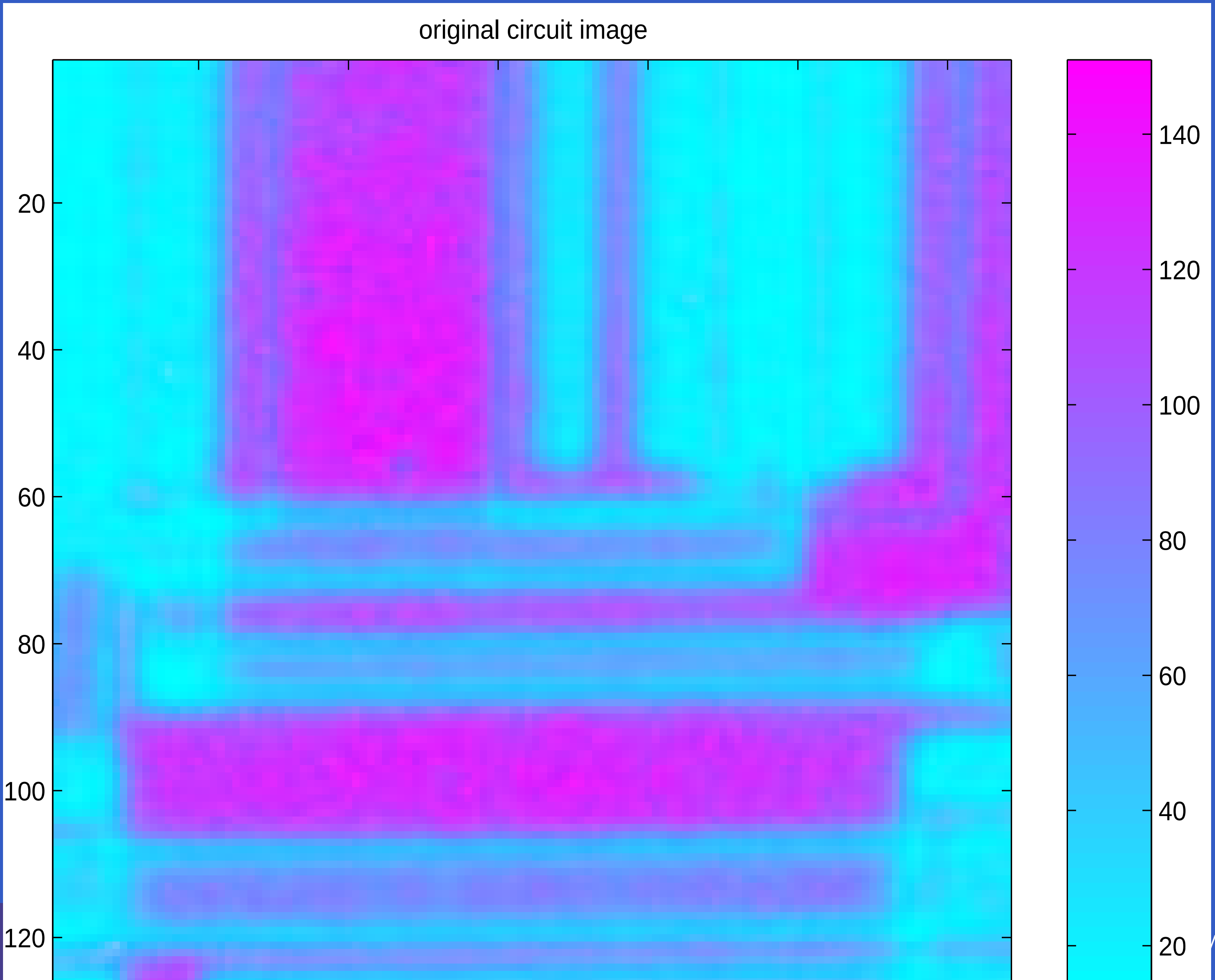
Use rank-revealing factorization $A \approx H\text{diag}(d)G$.

Examples :

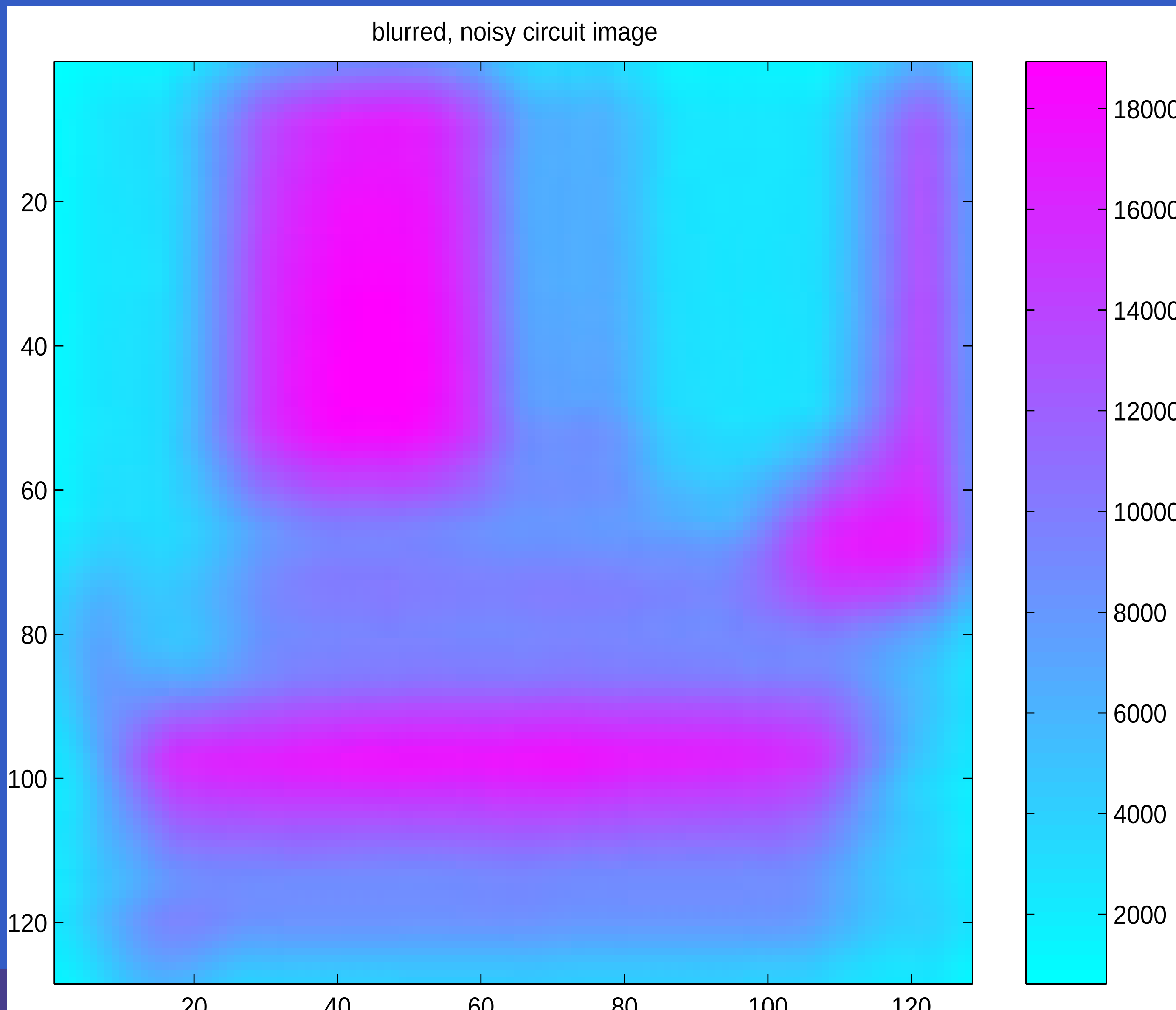
- Approx. by $A_1 \otimes A_2$ and use SVD [Kamm & Nagy '98]
- Circulant (BCCB) approximations [Hanke, Nagy, Plemmons '93]
- Fast, complete-pivoted LDU factorization of transformed matrix [K. & O'Leary '99, K. '99]

Example

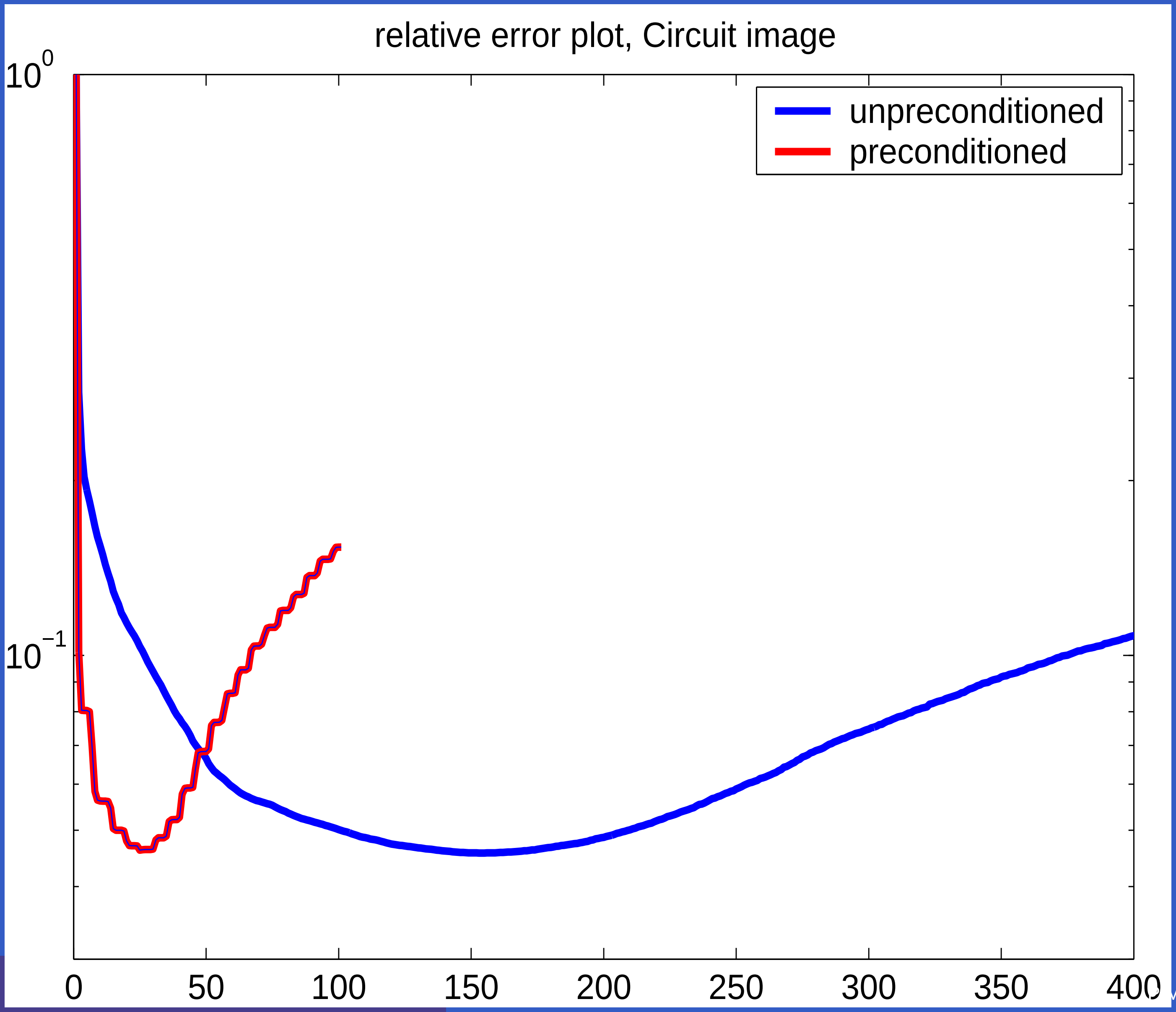
A is non-separable $128^2 \times 128^2$, $\approx .1$ % noise



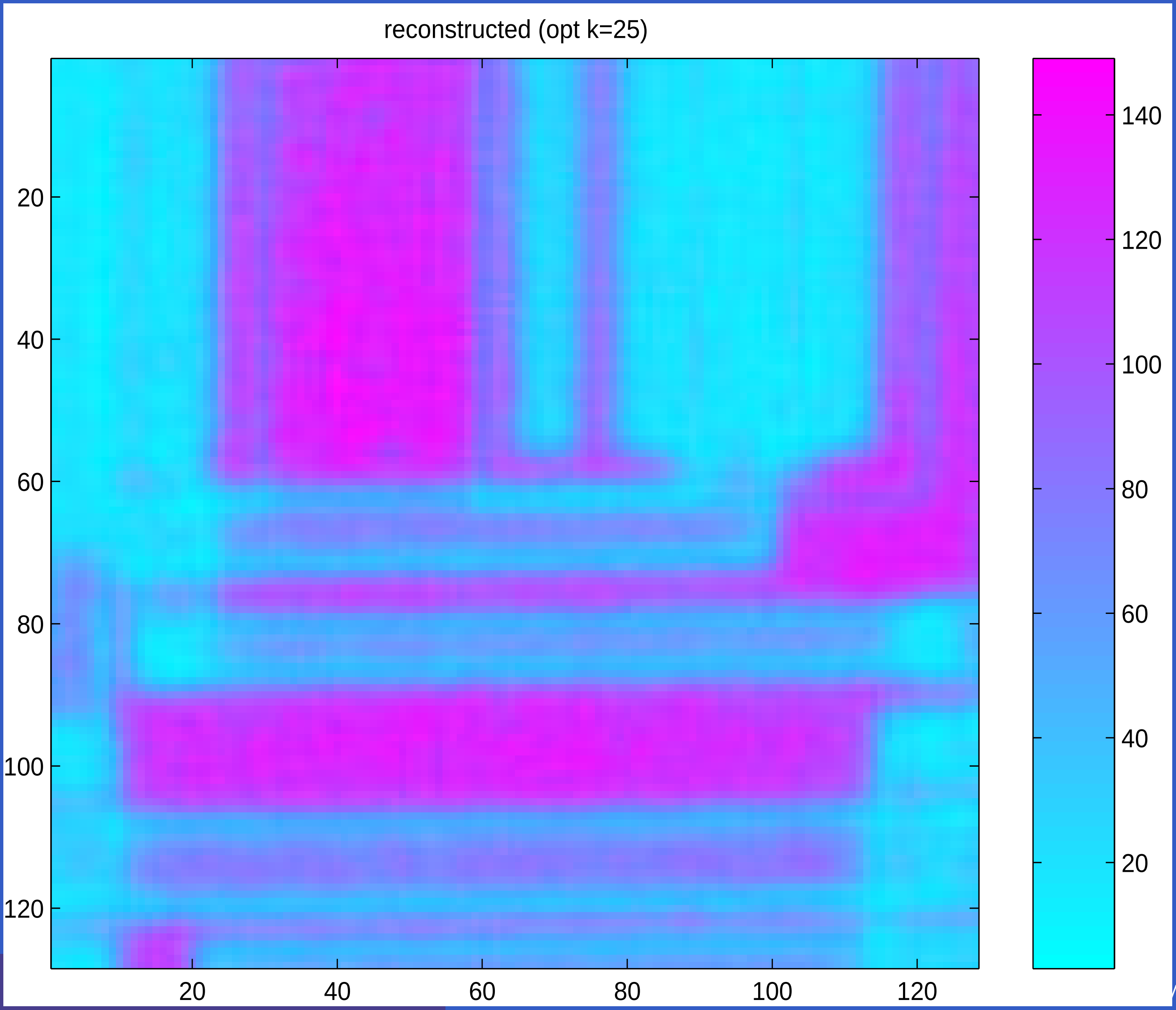
Example, cont



Relative Error Comparison



Restoration



Hybrid Approach (LSQR)

LSQR details:

$$AV_k = U_{k+1} B_k$$

where

- $V_k = [v_1, v_2, \ldots, v_k], U_k = [u_1, \ldots, u_k]$ have orthonormal columns
- B_k is $(k+1) \times k$ lower bidiagonal matrix.
- Columns of V_k span $\mathcal{K}_k(A^T A, A^T b)$

LSQR

Also, $u_1 = \beta e_1$. So

$$f^{(k)} = \arg \min_{z \in K_k(A^T A, A^T g)} \|Az - g\|_2$$

becomes

$$y^{(k)} = \arg\min_{y \in \mathbb{R}^k} \|B_k y - \beta e_1\|_2, \qquad f^{(k)} = V_k y_k.$$

The projected problem inherits properties of the original!

Hybrid Methods

The hybrid approach [O'Leary & Simmons '81]:

- Regularize the projected problem

  e.g. use TSVD to form $y_{reg}^{(k)}$

- The regularized solution is $f_{reg}^{(k)} = V_k y_{reg}^{(k)}$

Advantages of Hybrid Approach

Advantages:

- Cheap if k is small (preconditioning)
- Choose the regularization parameter using your favorite method [K. & O'Leary '01]
- Insensitivity to k

Tikhonov Regularization

Tikhonov regularized problem:

$$\min_f \left\{ \|Af - b\|_2^2 + \lambda^2 \|Lf\|_2^2 \right\}$$

or

$$\min_f \left\| \begin{bmatrix} A \\ \lambda L \end{bmatrix} f - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2 .$$

Conditioning depends on regularization parameter $\lambda > 0$.

Tikhonov Regularization and LSQR

In standard form Tikhonov, $L = I$.

$$f^{(k,\lambda)} = \arg\min_{\mathcal{K}_k} \left\| \begin{bmatrix} A \\ \lambda I \end{bmatrix} f - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2$$

where

$$\begin{aligned} \mathcal{K}_k &\equiv \mathcal{K}_k(A^T A + \lambda^2 I, A^T b) \\ &\equiv \mathcal{K}_k(A^T A, A^T b) \end{aligned}$$

Tikhonov and LSQR

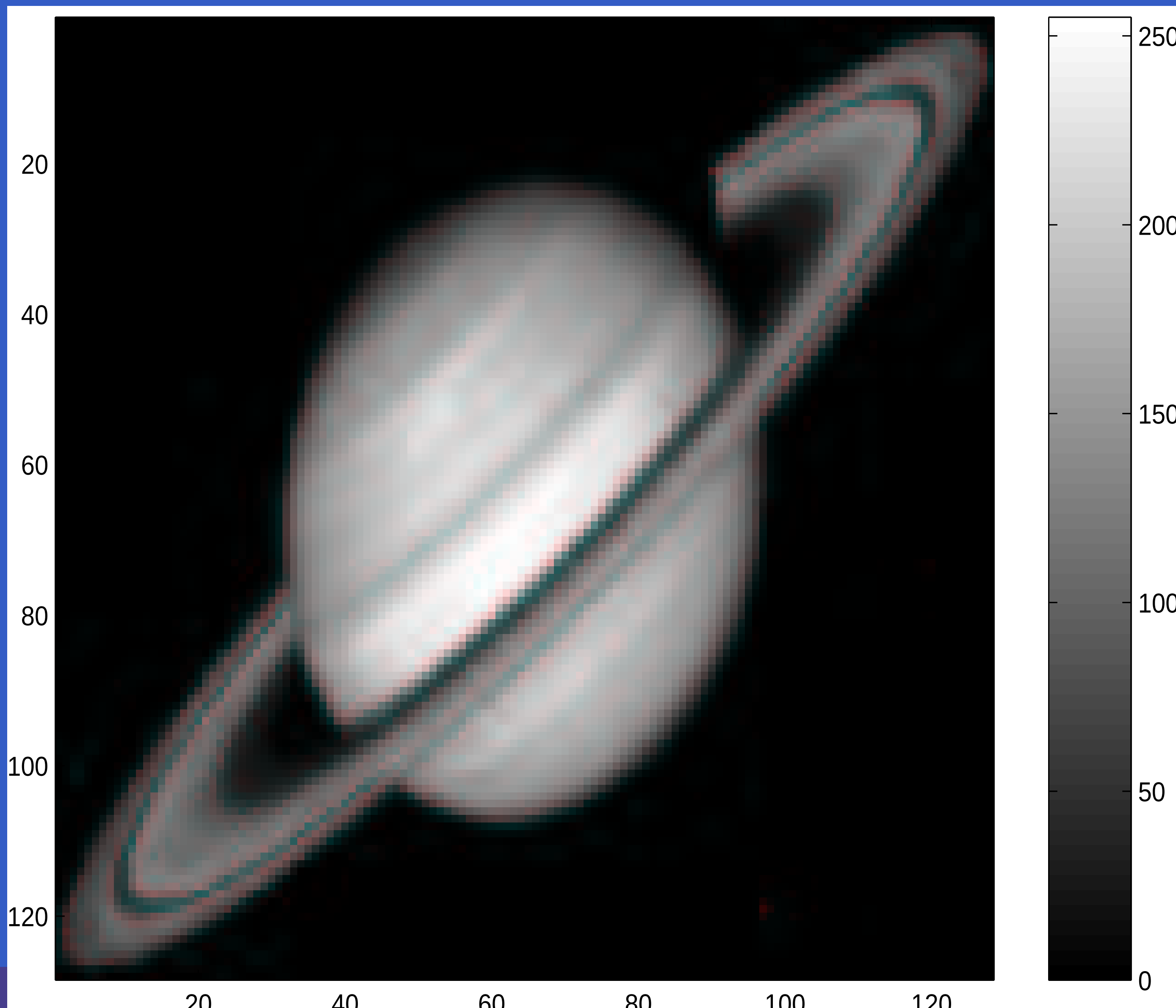
However

$$y^{(k,\lambda)} = \arg\min_{y \in \mathbb{R}^k} \left\| \begin{bmatrix} B_k \\ \lambda I \end{bmatrix} y - \begin{bmatrix} \beta e_1 \\ 0 \end{bmatrix} \right\|_2,$$

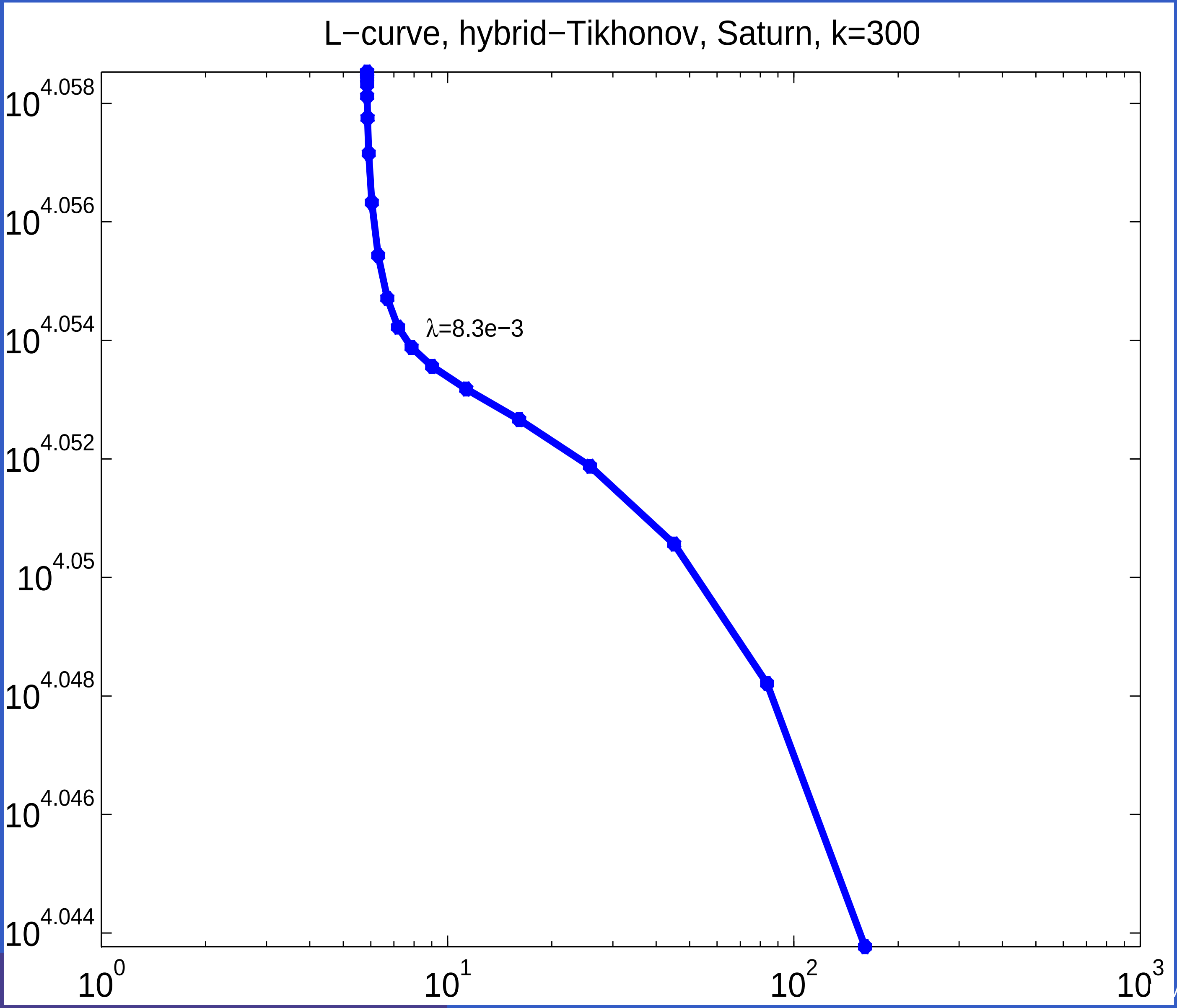
and $f^{(k,\lambda)} = V_k y^{(k,\lambda)}$

$\Longrightarrow$ Hybrid idea applicable here!

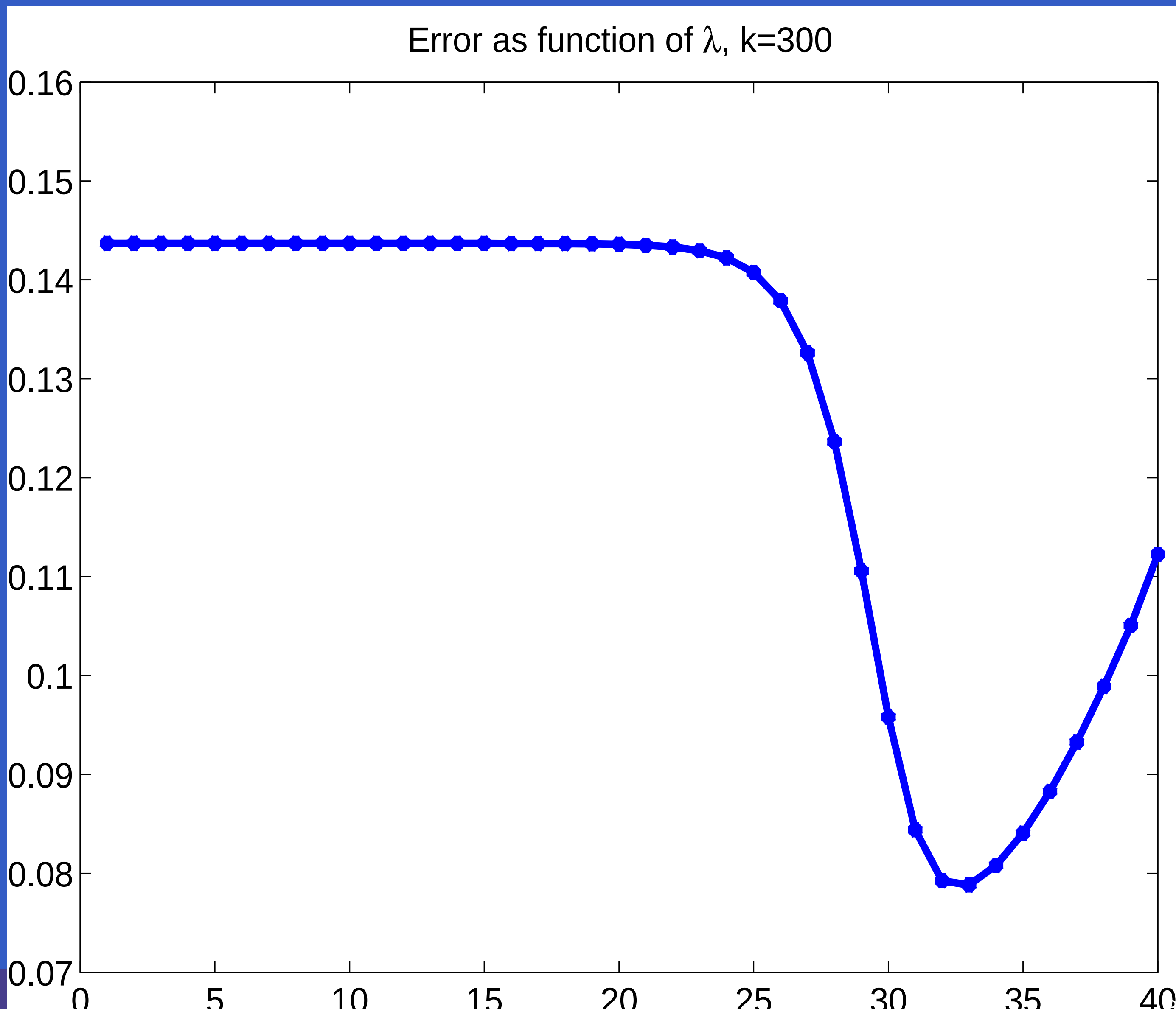
Example



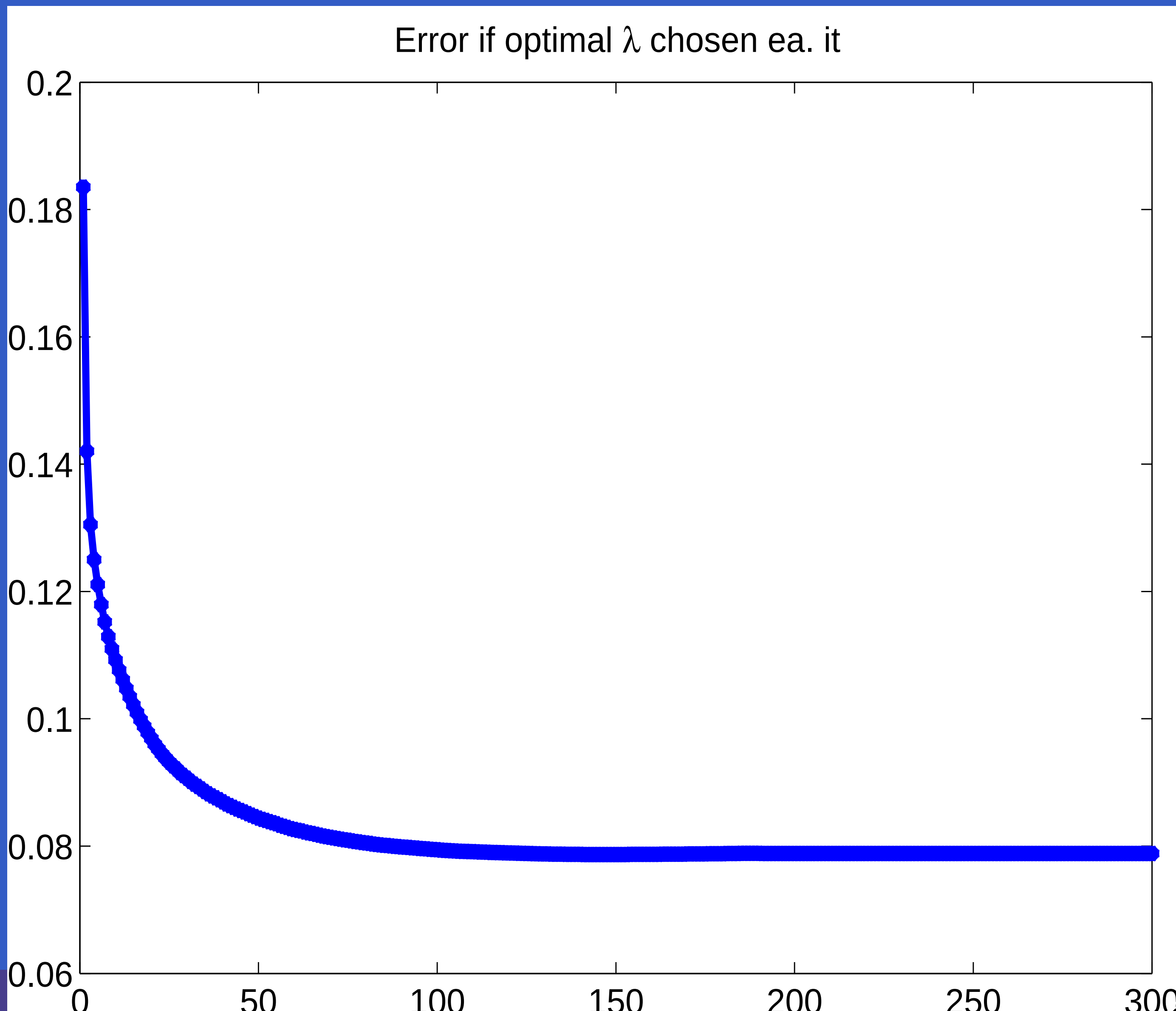
Saturn example



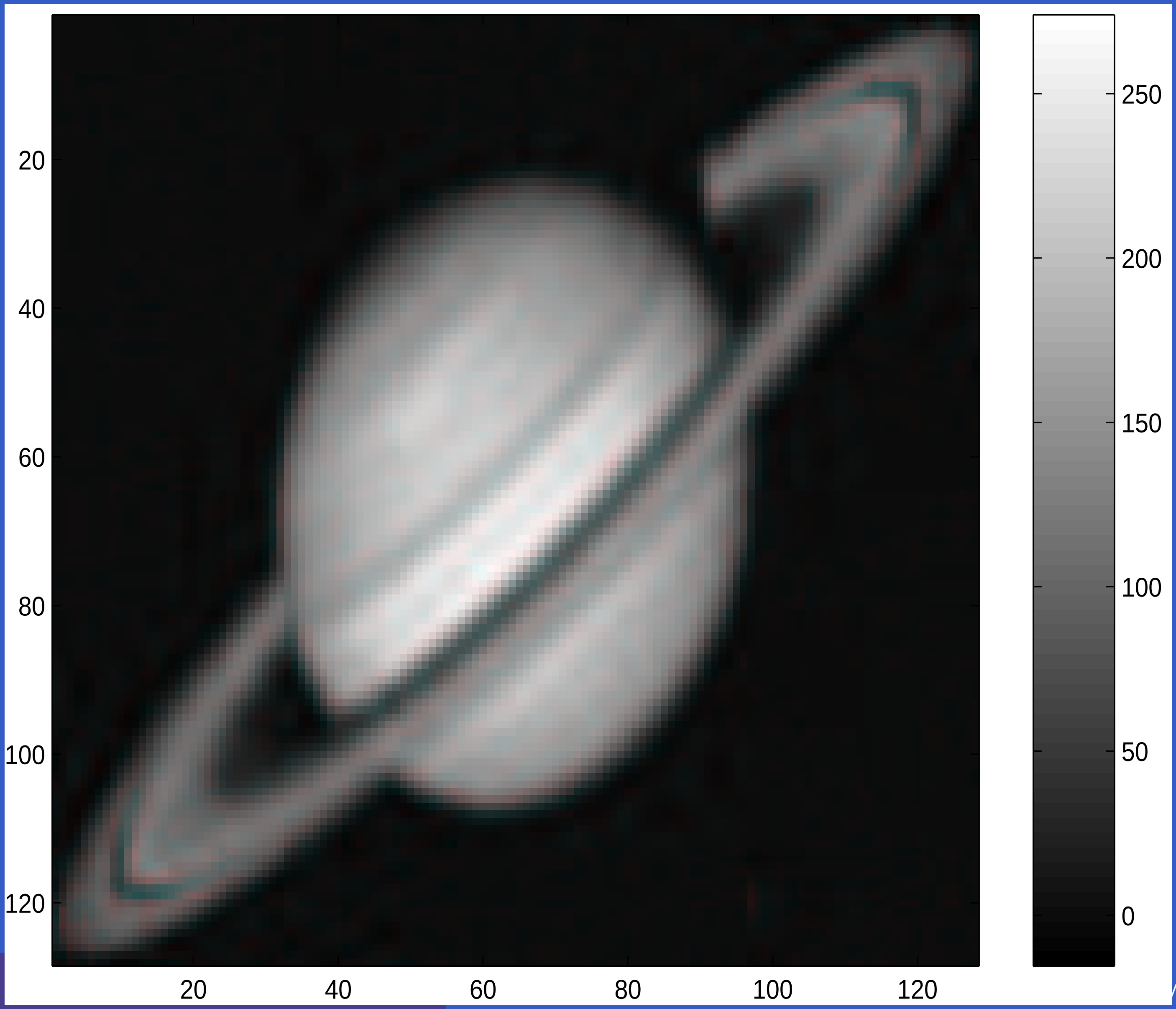
Saturn example



Saturn Example



Saturn Example



Tikhonov Regularization, $L \neq I$

- If λ is known, use LSQR to iteratively solve

$$\min_f \left\| \begin{bmatrix} A \\ \lambda L \end{bmatrix} f - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2 .$$

Preconditioning [Hanke & Vogel '99]

- Otherwise, transform to standard form:

$$\min_z \left\| \begin{bmatrix} AL_A^\dagger \\ \lambda I \end{bmatrix} z - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2 , \qquad f^{(\lambda)} = L_A^\dagger z^{(\lambda)}$$

Choosing λ

If iterative solver used, $L_A^\dagger$ is applied implicitly.

- λ computed by e.g. GCV, L-curve, on projected problem
- L-ribbon [Calvetti, Golub & Reichel '99], or
- curvature-ribbon [Calvetti, Hansen, & Reichel '02]

Alternatives

When

- λ not known a priori,
- computing/applying $L_A^\dagger$ not feasible,
- preconditioner for $\begin{bmatrix} A \\ L \end{bmatrix}$ known or not needed

use iterative LSQR-like approach that does not require transformation to standard form [Hansen, Jacobsen, & K. ’03]

Summary

- Krylov-subspace methods can be effective regularization methods
- Need to choose stopping parameter
- May need to precondition to be efficient
  - Cluster part of the spectrum
  - Do not mix signal and noise subspace
- Hybrid approaches an alternative
- Can include prior information with a Tikhonov approach