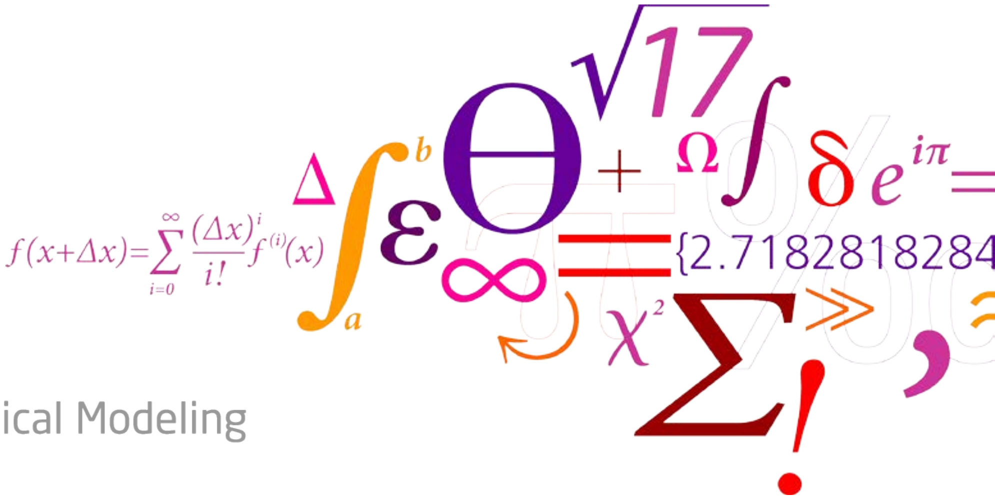
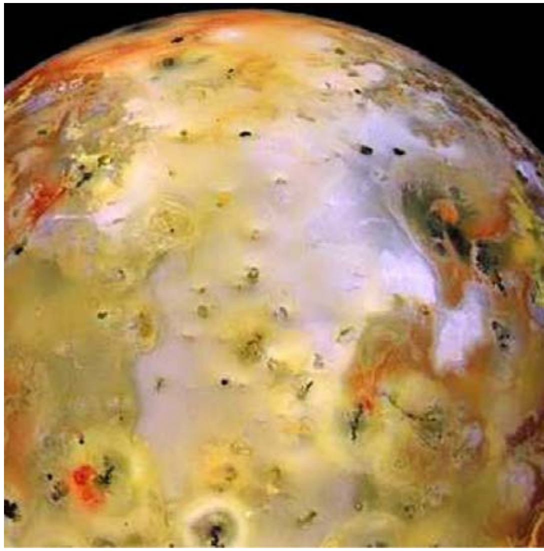
A Parameter-Choice Method That Exploits Residual Information

Per Christian Hansen
Section for Scientific Computing
DTU Informatics

Joint work with
Misha E. Kilmer
Tufts University

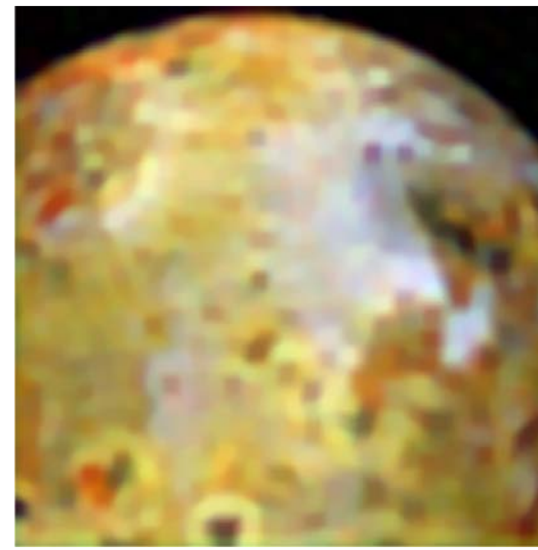


Inverse Problems: Image Deblurring



blurring

deblurring



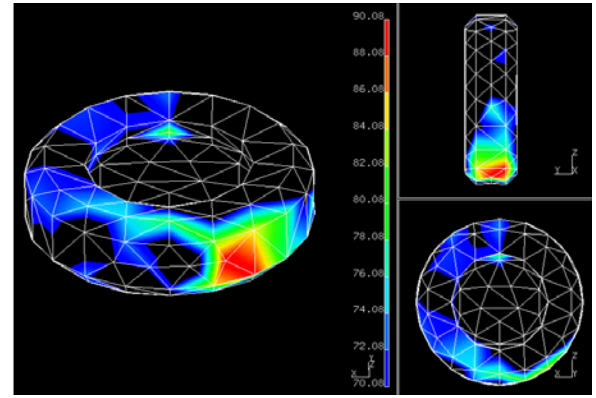
Io (moon of Jupiter)

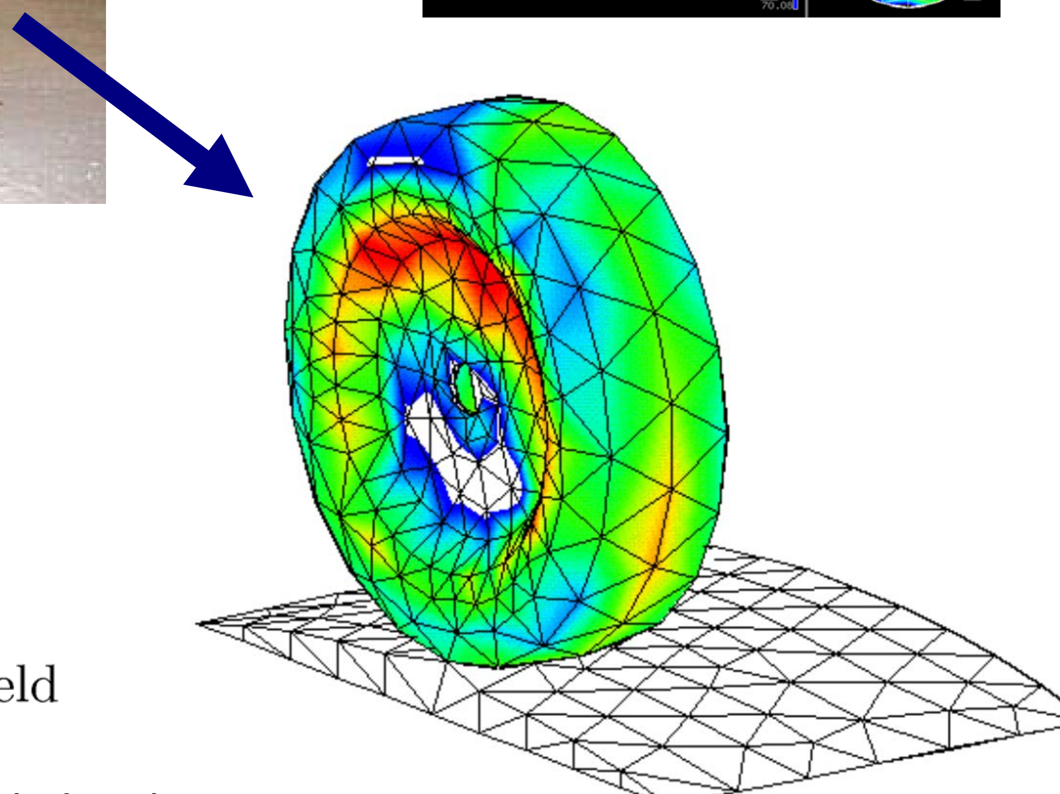
$$\int_{\Omega} K(\mathbf{s},\mathbf{t})\, f(\mathbf{t})\, d\Omega = g(\mathbf{s})$$

$$\begin{aligned} f(\mathbf{t}) &= \text{true scenery} \\ g(\mathbf{s}) &= \text{data (blurred image)} \\ K(\mathbf{s},\mathbf{t}) &= \text{point spread function} \end{aligned}$$

You cannot depend on your eyes when your imagination is out of focus
– *Mark Twain*

Sound Source Reconstruction

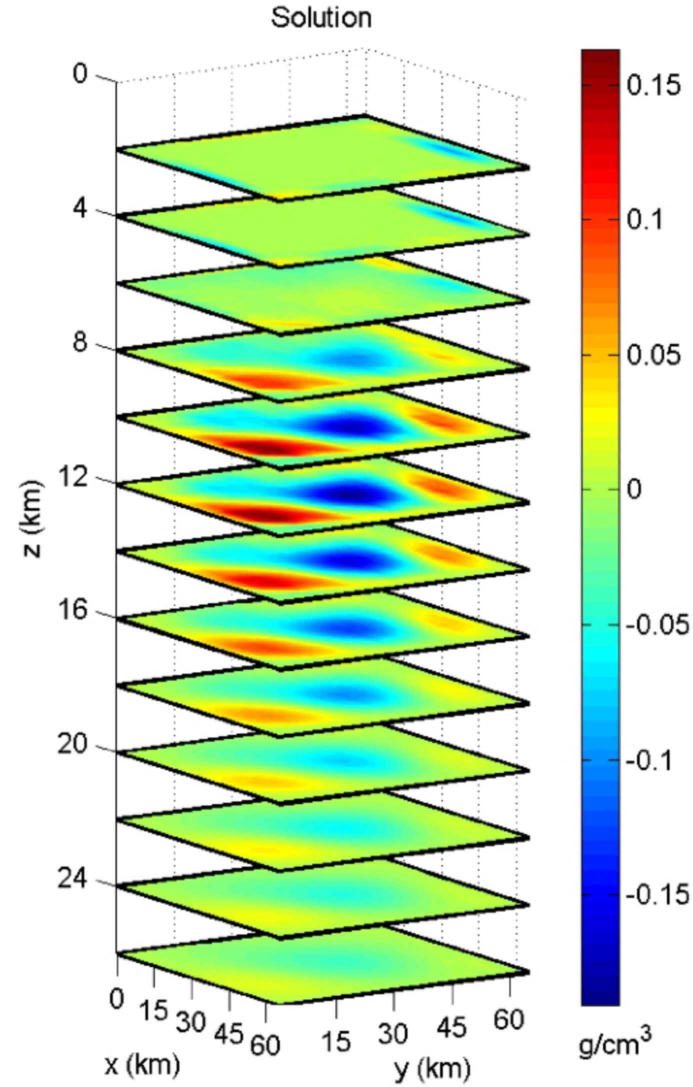




$$\int_\Omega K(\mathbf{s},\mathbf{t})\, f(\mathbf{t})\, d\Omega = g(\mathbf{s})$$

$$\begin{aligned} f(\mathbf{t}) &= \text{surface velocity} \\ g(\mathbf{s}) &= \text{data (pressure)} \\ K(\mathbf{s},\mathbf{t}) &= \text{acoustic dipole field} \end{aligned}$$

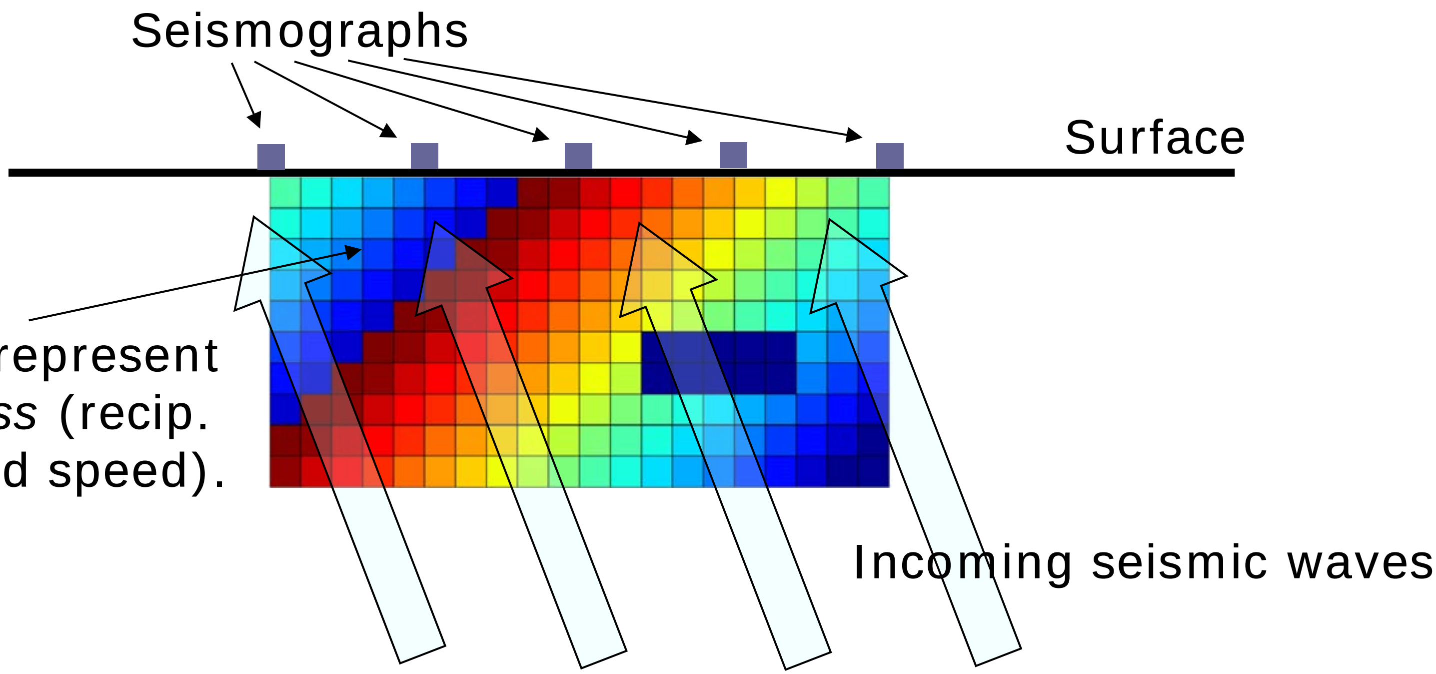
Potential Field Inversion



$$\int_\Omega K(\mathbf{s},\mathbf{t})\, f(\mathbf{t})\, d\Omega = g(\mathbf{s})$$

$$\begin{aligned} f(\mathbf{t}) &= \text{magnetization} \\ g(\mathbf{s}) &= \text{data (anomaly)} \\ K(\mathbf{s},\mathbf{t}) &= \text{magnetic dipole field} \end{aligned}$$

Seismic Tomography

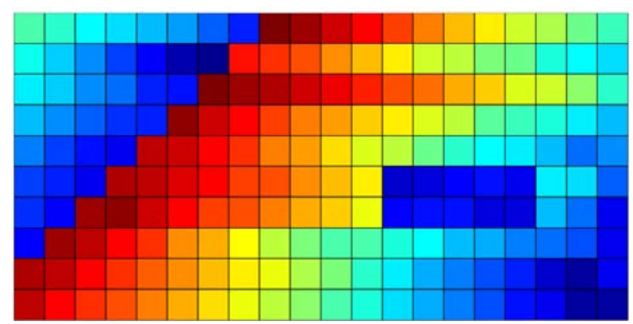


The line integrals

$$b_i = \int_{\mathrm{ray}_i} f(\tau)\, d\tau$$

give the relationship between the material property f and the measured travel times b_i.

Reconstruction



Inverse Problems

Goal: find the (hidden) *source* that gives rise to the measured *data* through a *model* for the source's action.

Inverse problems are examples of ill-posed problems:

- the solution may not exist,
- the solution may not be unique, or
- the solution may not depend continuously on data.

The linear systems of equations $A\,x = b$ associated with our discretizations are always **ill conditioned!**

Consequence: solutions are extremely sensitive to errors in our data!!

SVD Analysis

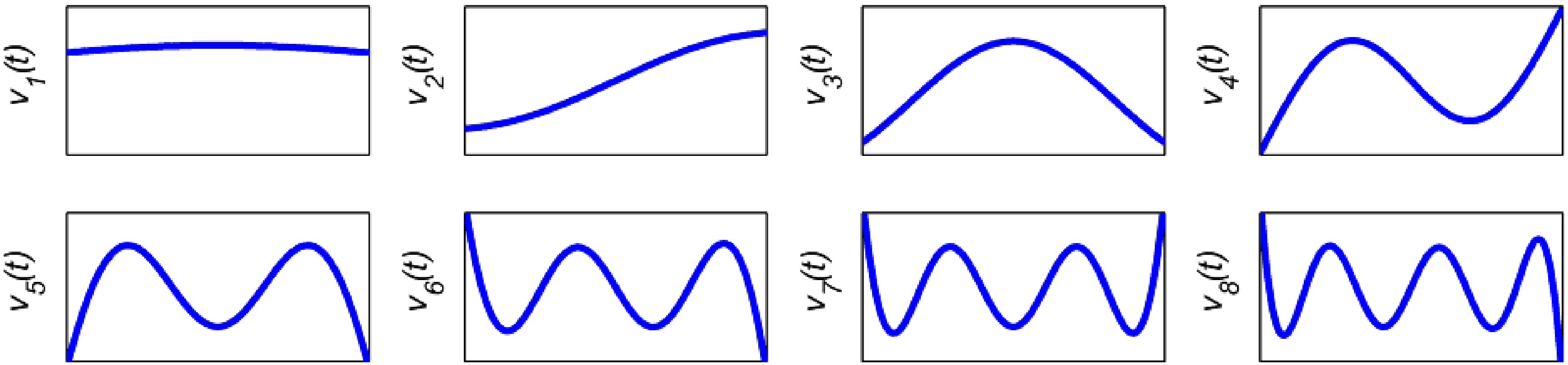
The SVD: $A = U \Sigma V^T, \quad U^T U = V^T V = I, \quad \Sigma = \mathrm{diag}(\sigma_i).$

The "naive solution" is

$$A^{-1} b = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i .$$

We can not use $A^{-1}b$ because $u_i^T b / \sigma_i \rightarrow$ LARGE for $i \rightarrow n$.

The singular vectors looks like a spectral basis (we'll return to this):



Regularization

We must apply regularization in order to deal with the ill conditioning of the problem and suppress the influence of the noise in the data.

Tikhonov regularization:

$$\min \left\{ \|A\,x - b\|_2^2 + \lambda^2 \, \|x\|_2^2 \right\} \quad \Rightarrow \quad x_\lambda = A_\lambda^\dagger b.$$

The choice of smoothing norm and parameter λ makes the problem regular and ensures that we compute a robust solution x_λ.

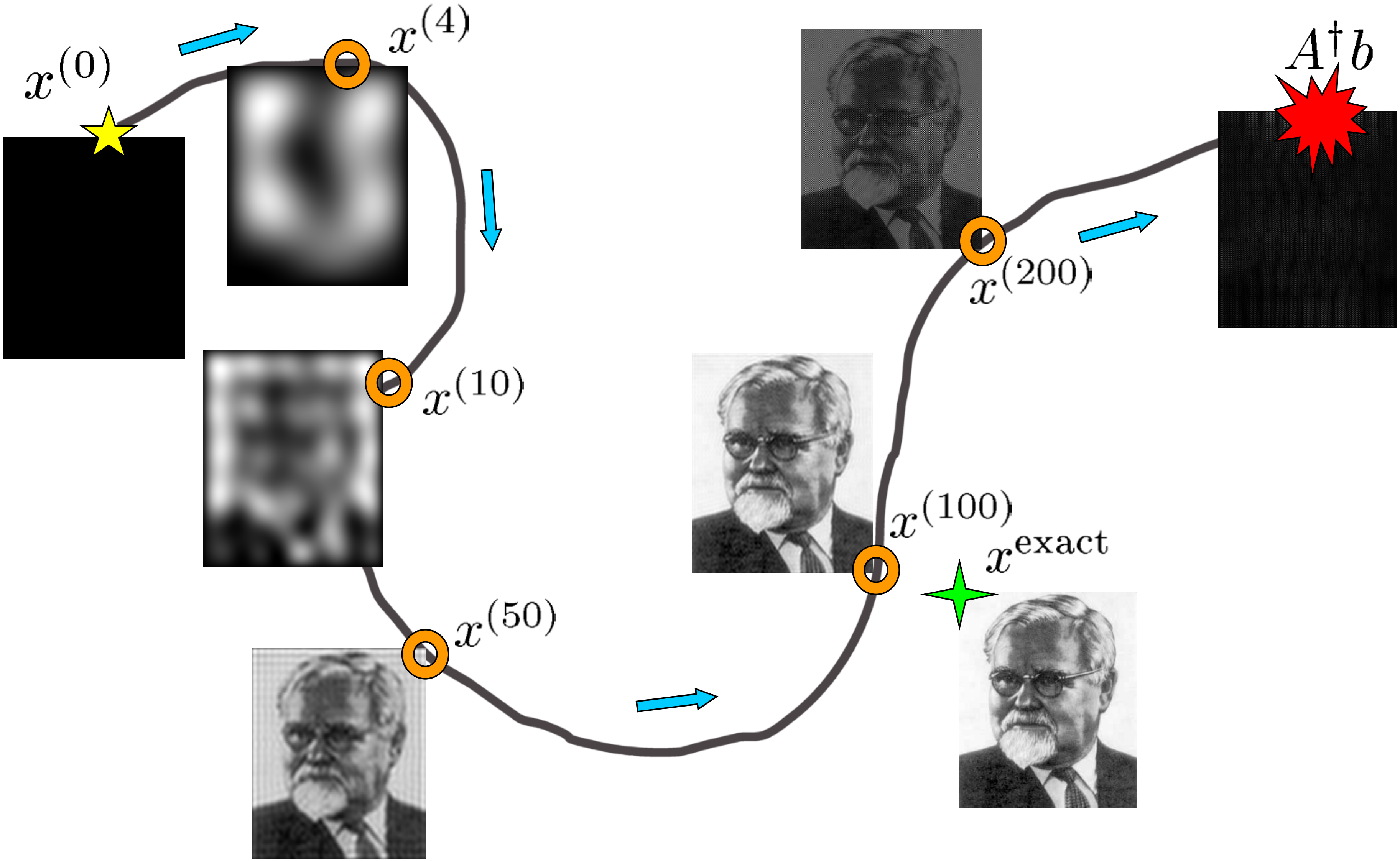
Regularization by projection:

$$\min \|A\,x - b\|_2 \quad \text{subject to} \quad x \in \mathcal{S}_k$$

where $\mathcal{S}_k$ is a k-dimensional subspace spanned by desirable basis vectors:

- TSVD: $\mathcal{S}_k = \text{span}\{v_1, v_2, \ldots, v_k\}$,
- CGLS: $\mathcal{S}_k = \text{span}\{A^T b\,,\ A^T A\, A^T b\,,\ (A^T A)^2 A^T b\,,\ \ldots\}$.

Semi-Convergence of CGLS

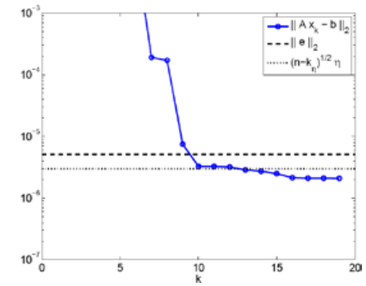


The Quest for the Holy Grail = λ (or *k*)

Many of the current parameter-choice methods are based on *norms*.

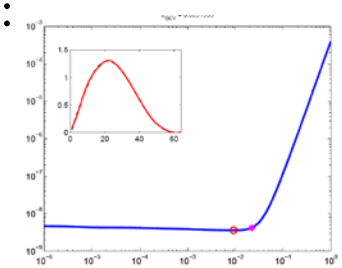
Discrepancy Principle. Fit to the error level in the data:

$$\text{find } \lambda \text{ so } \|A\,x_\lambda - b\|_2 \approx \|e\|_2 .$$



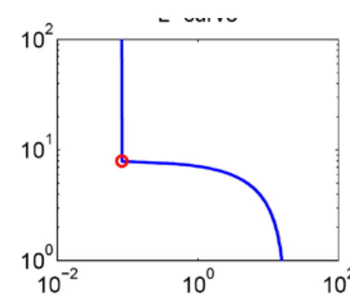
Generalized Cross Validation. Minimize prediction error:

$$\text{minimize } \frac{\|A\,x_\lambda - b\|_2}{\operatorname{trace}(I - A A_\lambda^\dagger)} \quad .$$



L-Curve Criterion. Find the "corner" of the curve

$$(\,\log \|A\,x_\lambda - b\|_2 \,,\ \log \|L\,x_\lambda\|_2\,).$$



The Three Golden Rules of Inversion

According to Bill Lionheart, Manchester University (augmented by PCH):

1. Understand the measured data and their errors.
2. Be precise about what you want from the solution.
3. Incorporate what you already know about the solution.
4. Understand the forward model, incl. model errors.
5. Don't expect mathematics to compensate for lack of knowledge of the above.
6. There is no parameter-choice rule that will work for all problems!

A New Approach

The original idea is due to Bert Rust (Comp. Sci. Stat. 2000):

When the true solution is not known, the residual vector provides the only objective guide for assessing the quality of an estimate.

The strategy is to go *beyond norms:*

- Make use of more information present in the residual vector.
- Main ingredients of our analysis:
  - SVD/Fourier analysis,
  - statistical NCP analysis.

Our goal:

- Develop a statistically-motivated parameter selection method.
- Extract precisely all relevant "information" from the data.

The Discrete Fourier Transform (DFT)

The DFT is often written in "$\sum_{i=0}^{n-1}$ notation."

Following Van Loan, we prefer a matrix notation where the DFT is represented by a unitary matrix F such that:

- 1D signal $x \in \mathbb{R}^n$:

$$\texttt{dft}(x) = F^H x$$

- 2D signal $X \in \mathbb{R}^{n \times n}$:

$$\texttt{dft2}(X) = F^H X \, (F^H)^T = F^H X \, \texttt{conj}(F)$$

with

$$F_{jk} = \frac{1}{\sqrt{n}} \exp(2\pi\sqrt{-1}(j-1)(k-1)/n), \qquad i, j = 1, \ldots, n.$$

The use of the factor $1/\sqrt{n}$ is *non-standard notation*, but it makes the similarities with SVD analysis much clearer.

The 1D Normalized Cumulative Periodogram

Real signals have complex symmetric spectra, and we need only

$$F(:,1:q)^H x, \qquad q = \lfloor n/2 \rfloor + 1.$$

The periodogram for a 1D signal $x \in \mathbb{R}^n$ is the vector

$$p = |F(:,1:q)^H x|^2 \in \mathbb{R}^q$$

(called the power spectrum in EE).

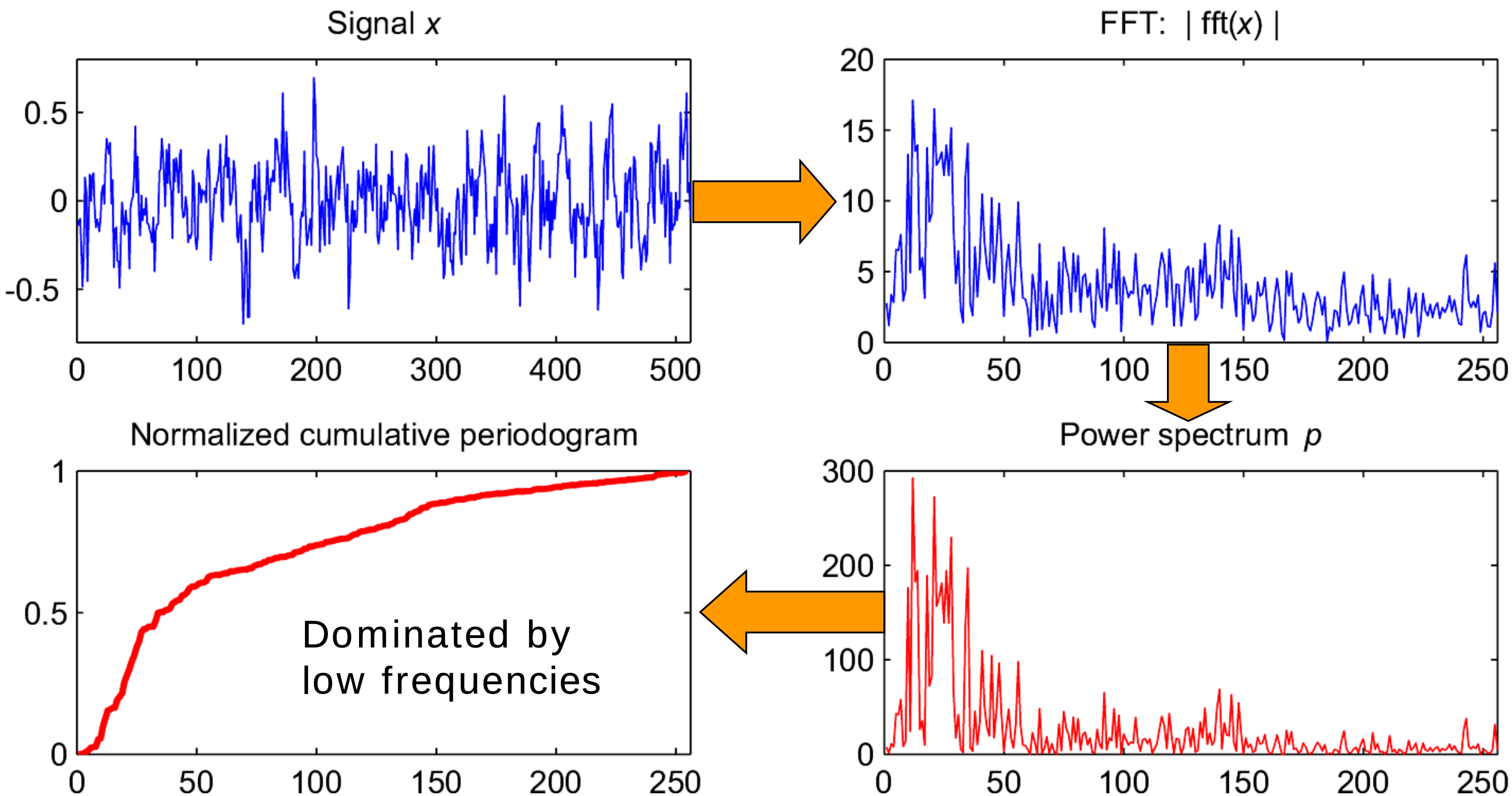
The *Normalized Cumulative Periodogram* (NCP) for the signal x is given by the vector $c(x) \in \mathbb{R}^{q-1}$ with elements

$$c(x)_k = \frac{\|p(2:k+1)\|_1}{\|p(2:q)\|_1}, \qquad k = 1, \ldots, q-1.$$

The "DC component" $p(1)$ does not take part of the NCP.

Example of NCP Analysis

Sound signal x = ”big storm” (from sound database).

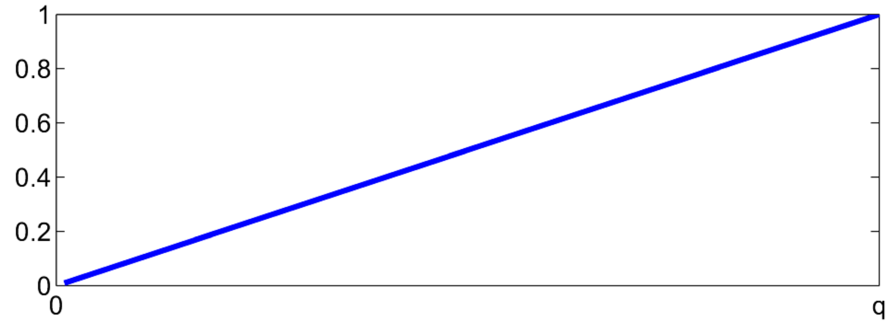


The NCP Reveals White Noise

White noise is invariant to orthogonal and unitary transformations:

$$\mathrm{Cov}(e) = \eta^2 I \quad \Rightarrow \quad \mathrm{Cov}(F^H e) = F^H \mathrm{Cov}(e) F = \eta^2 I.$$

Thus, the expected power spectrum for e is flat, and the expected NCP for e lies on a straight line between (0,0) and (q,1).

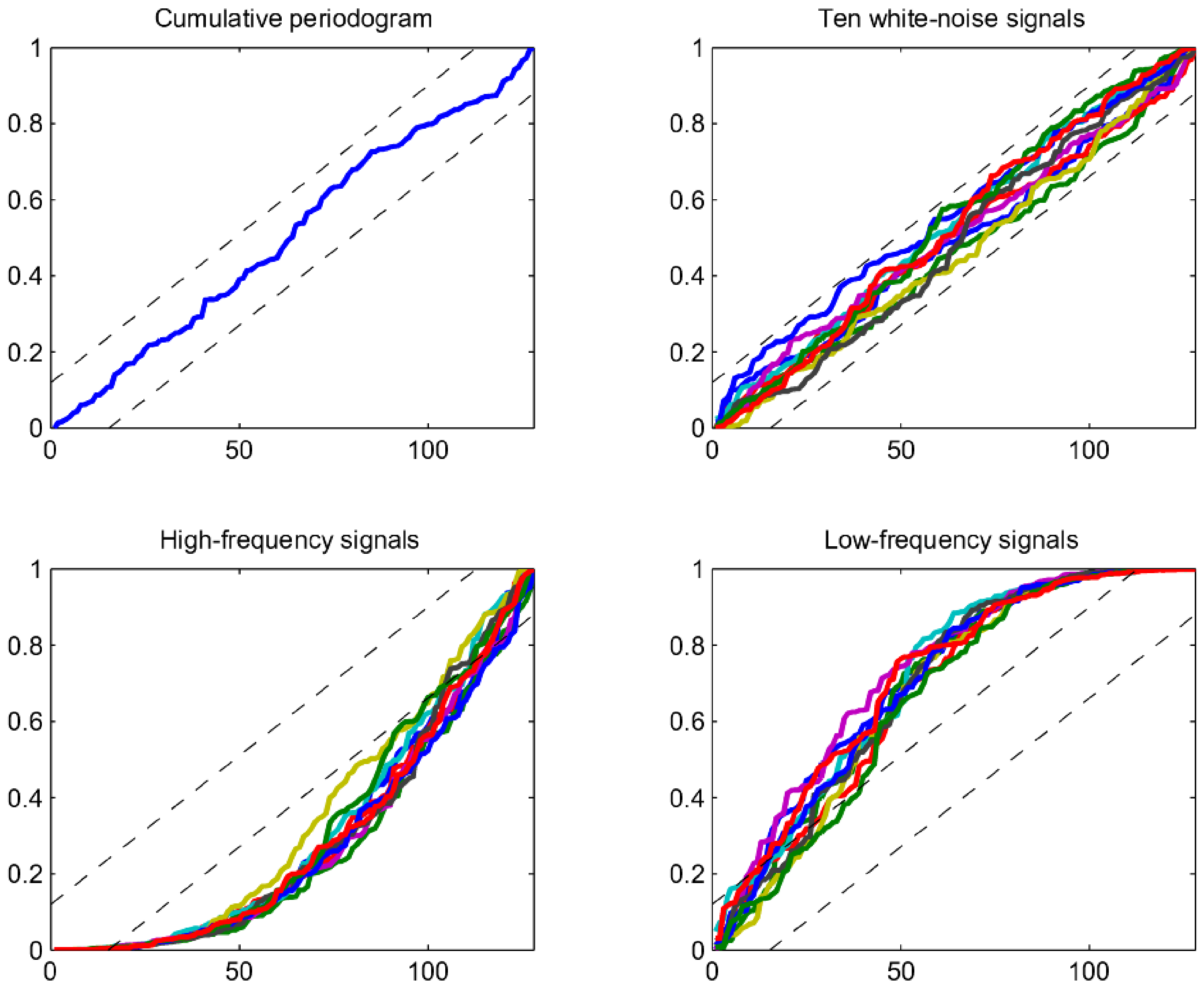


The NCP of a realization of the white-noise vector e should lie within the *Kolmogorov-Smirnoff limits* of this straight line.

For 5 percent significance level, the K-S limits are

$$\text{1D: } \pm 1.36\, q^{-1/2} \qquad \text{2D: } \pm 1.36\, q^{-1}$$

Examples of NCPs with K-S Limits



The 2D NCP

Don't read this – it's only here to show that we can treat 2D!

Define the matrix $P \in \mathbb{R}^{q \times q}$ (the 2D power spectrum) by

$$P = |F_{\triangleright}^{H} X \,\mathrm{conj}(F_{\triangleright})|^2.$$

Let Π be a permutation matrix such that the elements of the vector

$$\hat{p} = \Pi \,\mathrm{vec}(P) \in \mathbb{R}^{q^2}$$

are ordered in increasing spatial frequency.

Then the *2D NCP* for X is vector $c(X)$ with entries

$$c(X)_k = \frac{\|\hat{p}(2:k+1)\|_1}{\|\hat{p}(2:q^2)\|_1}, \quad k = 1, \ldots, q^2 - 1.$$

This extension to 2D problems is (to our knowledge) new.

Properties of the Residual

Recall that the residual vector for x_λ is

$$r_\lambda = b - A\,x_\lambda = U(I - A\,A_\lambda^\dagger)\,U^T b.$$

Key idea. Choose λ so that all pure signal has been removed from r_λ and only noise is left in the residual; design a test for this transition.

Recall that

$$\mathrm{Cov}(F^H e) = \eta^2 I$$

while

$$\mathrm{Cov}(F^H r_\lambda) = \eta^2 F^H U (I - A\,A_\lambda^\dagger)^2 U^T F.$$

Notice the "mixing" of Fourier and SVD coefficients due to $F^H U$.

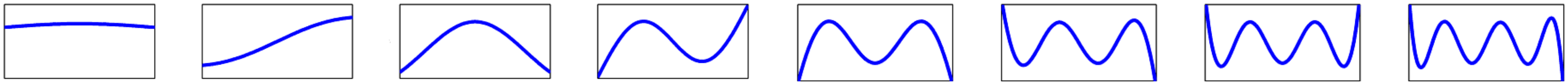
What can we say about the matrix $F^H U$ that will allow us to liken the Fourier analysis and the spectral filtering methods?

The Fourier Transform of the Residual

We skip technical details here – see the paper:

> P. C. Hansen, M. Kilmer, and R. H. Kjeldsen, *Exploiting residual information in the parameter choice for discrete ill-posed problems*, BIT, 46 (2006), pp. 41–59.

The singular functions share characteristics with the Fourier basis
$\rightarrow$ increasing oscillations as the singular values decrease.



The result is that the covariance matrix for the residual vector is

$$\mathrm{Cov}(F^H r_\lambda) \approx \eta^2 \begin{pmatrix} I_{\frac{m-k}{2}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_{\frac{m-k}{2}} \end{pmatrix}.$$

Unless k is large, the residual's noise component is white-noise like.

The New Parameter-Choice Rule

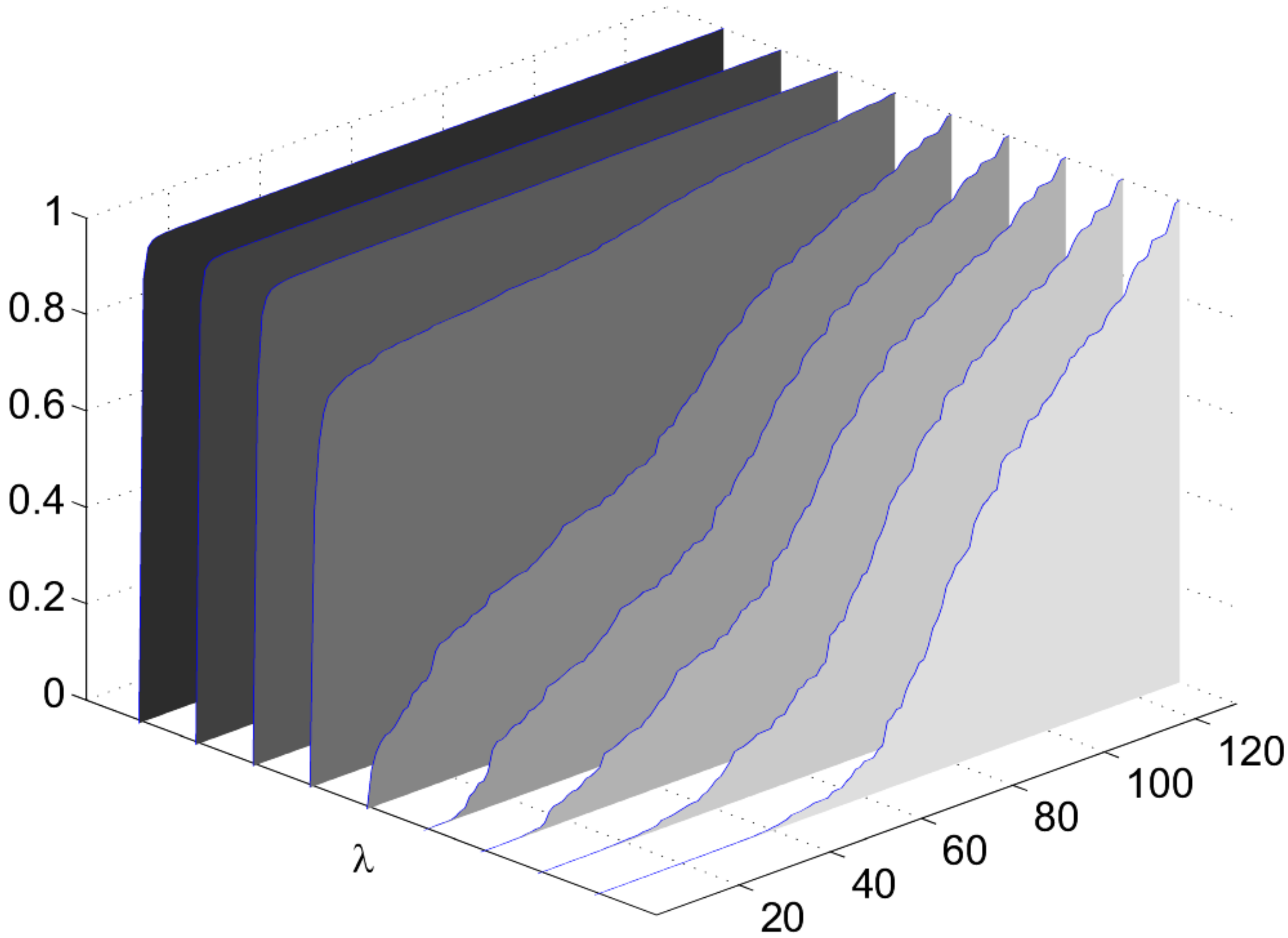
Choose the regularization parameter for which the residual vector *transitions* from being dominated by signal to white-noise like.

1. We have shown that this transition occurs in the SVD basis, when all relevant SVD components have been extracted.

2. Due to the similarity between the SVD and Fourier bases, we can use tests based on *Fourier analysis* and *NCP*.

3. Computational advantage: the FFT is fast!

NCPs for `deriv2` Test Problem



NCPs for `phillips` Test Problem

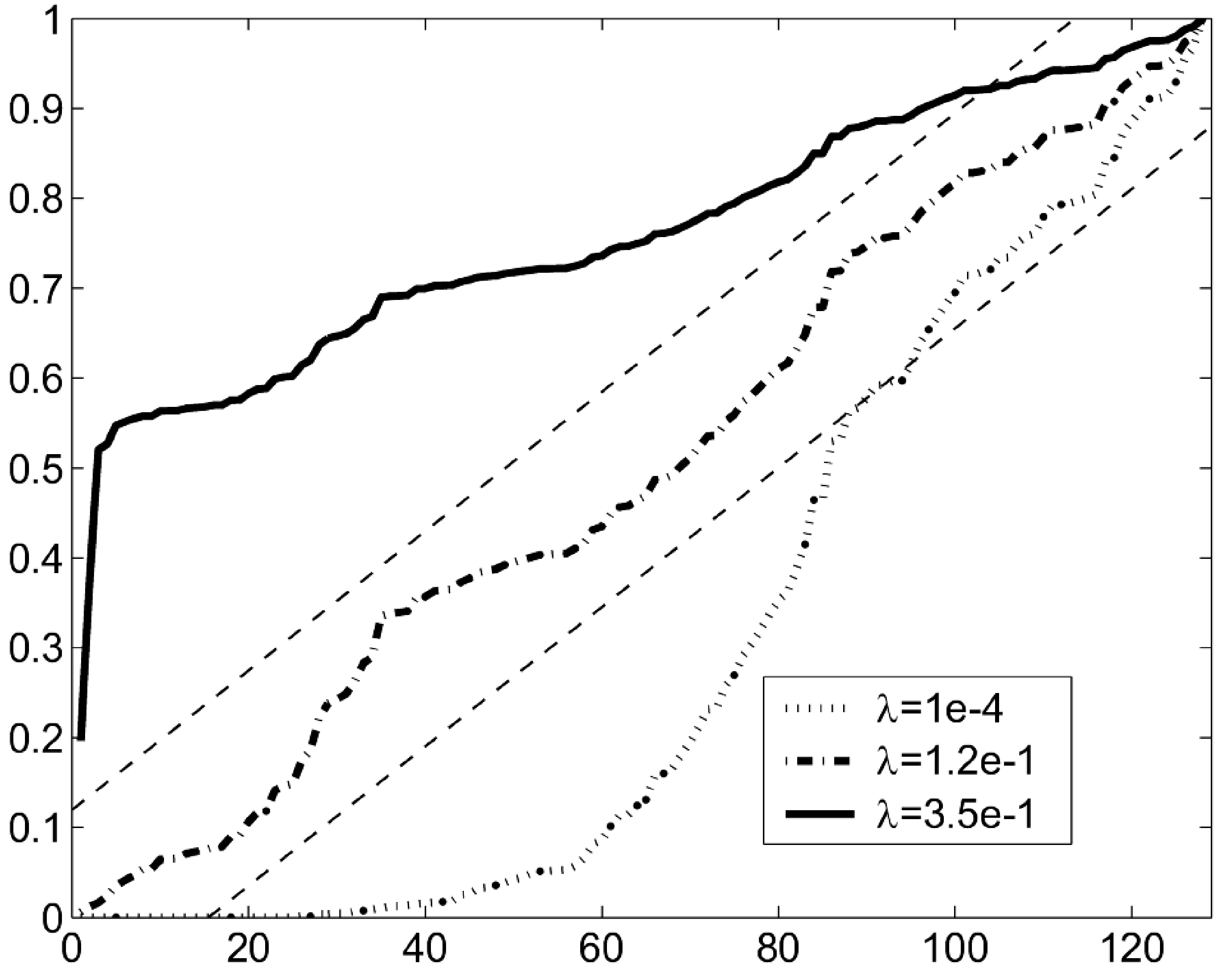
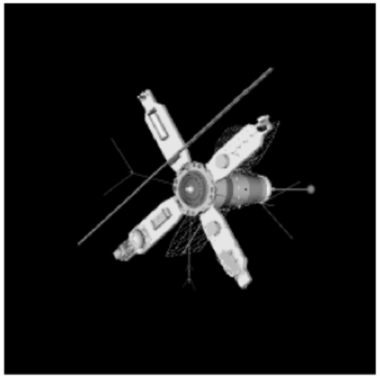


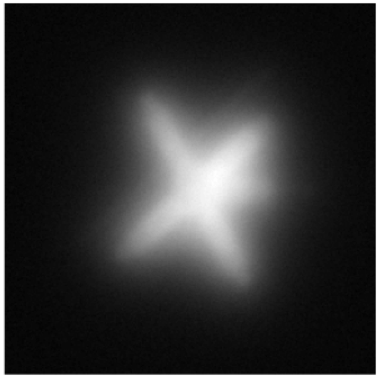
Image Deblurring Example

Noise level $\|e\|_2/\|\bar{b}\|_2 = 3 \cdot 10^{-2}$. CGLS – regularizing iterations.

Exact image



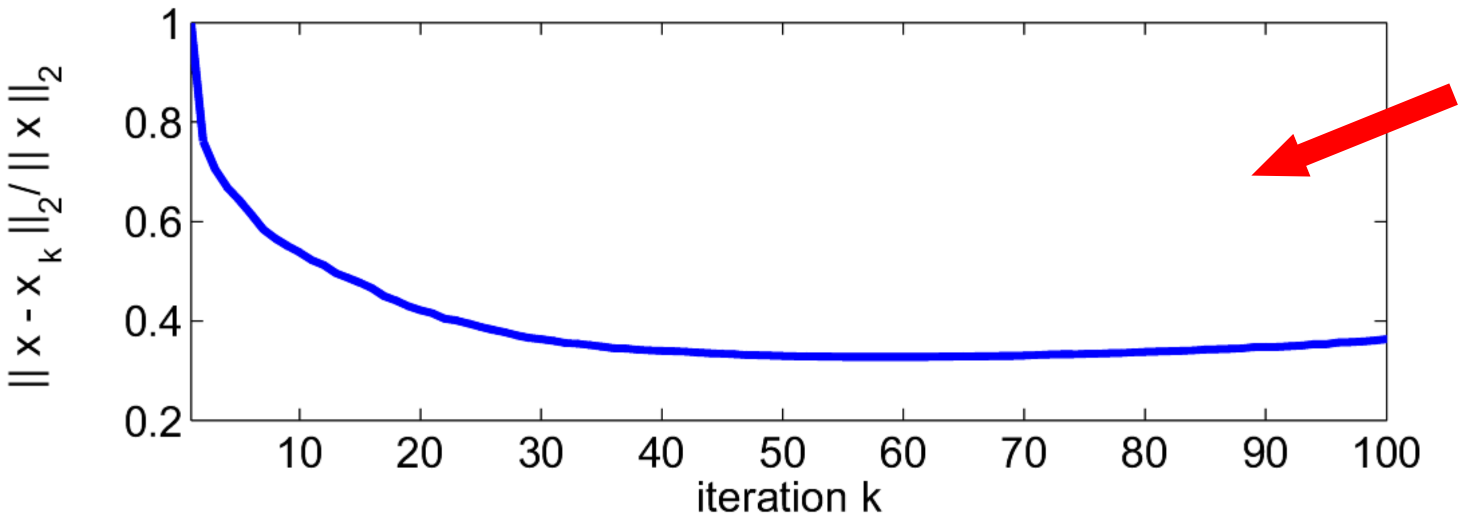
Blurred noisy image



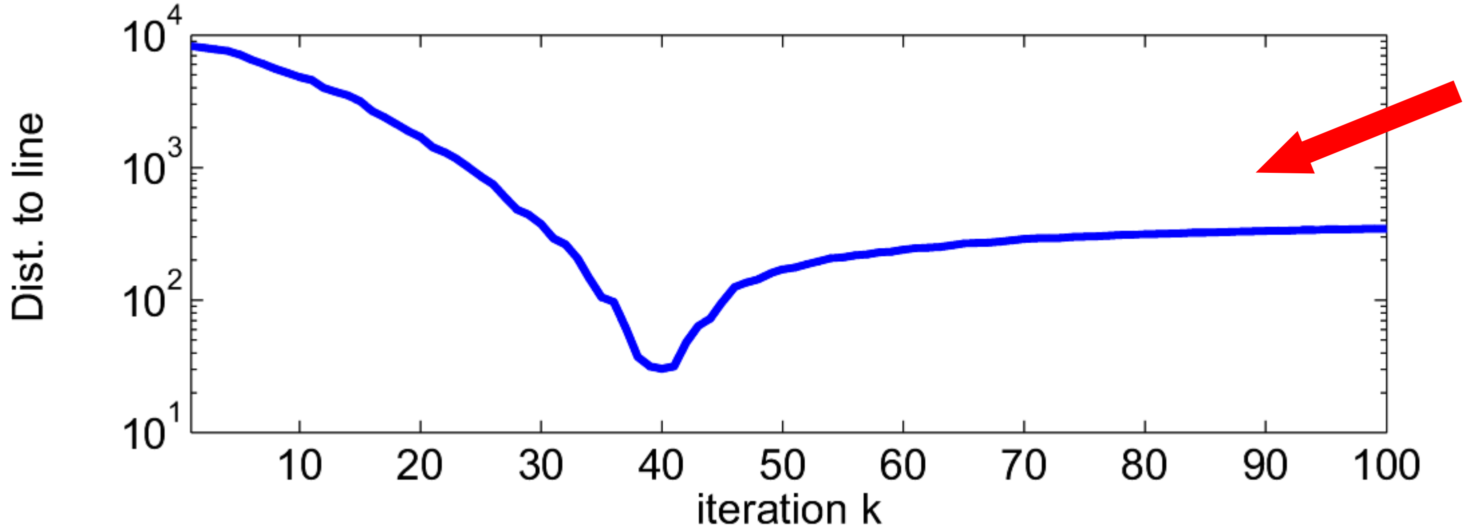
”Optimal” solution

NCP solution

Does It Work?



Error history



Distance from NCP to straight line

3D Tomography in Crystallography

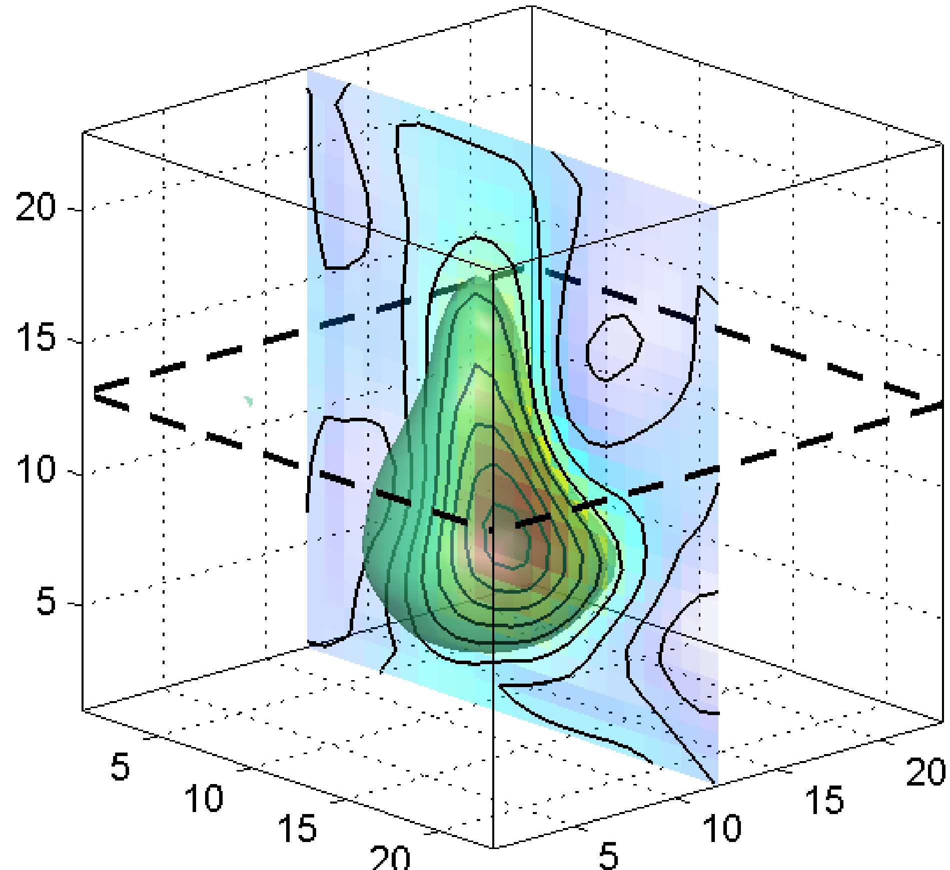
Joint work with **Metals in 4D**, Risø DTU, Denmark.

Single-ODF reconstruction:

- Data: X-ray diffraction
- Reconstruction:orientation distribution function (ODF)
- Smoothing norm: $|| \nabla^2 f ||^2$

Reconstruction method:

- CGLS regularizing iterations
- NCP stopping criterion.



Solution shows distribution of orientations in imperfect crystal.

Final Comments

- Our paper demonstrated the often-observed *relationship* between Fourier and SVD bases.
- Our study gives *insight* into how to exploit Fourier components of the residual.
- Our insight leads to convenient *parameter-choice rule* based on the FFT and the NCP.
- Related transforms can be used, such as the DCT.
- Can also be used in the presence of (low frequent) signal-correlated noise (see our paper).
- Sets the stage for other methods based on statistical analysis of residuals.

Thank you!