# Inverse Problems

## *Do the Impossible – Solve the Unsolvable*

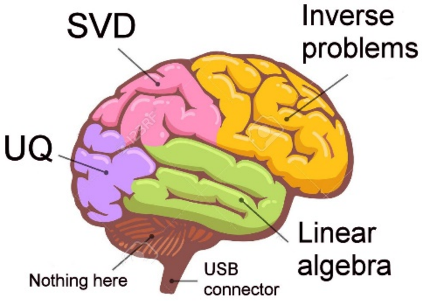
*Per Christian Hansen*
Professor, Villum Investigator
Section for Scientific Computing

What it’s like to do research

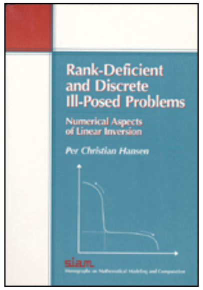
Heian Shrine, Kyoto

**DTU Compute**
Department of Applied Mathematics and Computer Science

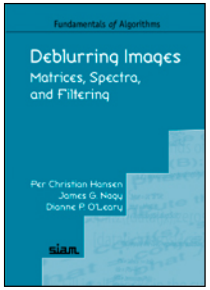
# About Me …



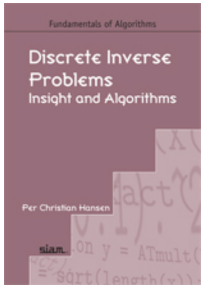
- Numerical analysis & inverse problems – regularization algorithms, matrix computations, image deblurring, signal processing, Matlab software, …
- Head of the Villum Investigator project
  Computational Uncertainty Quantification for Inverse Problems. ⇨ **CUQI**
- Author of several Matlab software packages.
- Author of four books (one more underway).



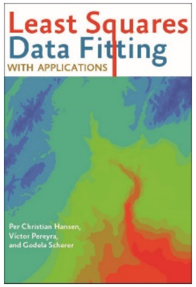
Rank-Deficient and Discrete Ill-Posed Problems
Numerical Aspects of Linear Inversion
Per Christian Hansen
siam



Fundamentals of Algorithms
Deblurring Images
Matrices, Spectra, and Filtering
Per Christian Hansen
James G. Nagy
Dianne P. O'Leary
siam



Fundamentals of Algorithms
Discrete Inverse Problems
Insight and Algorithms
Per Christian Hansen
siam



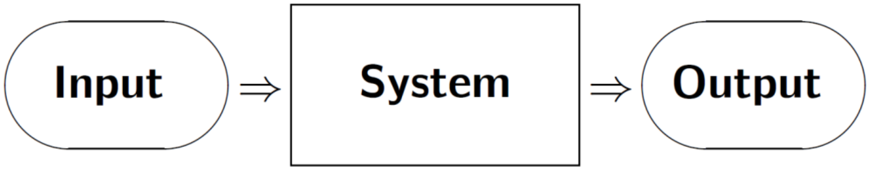
Least Squares Data Fitting
WITH APPLICATIONS
Per Christian Hansen, Victor Pereyra, and Godela Scherer

Computed Tomography

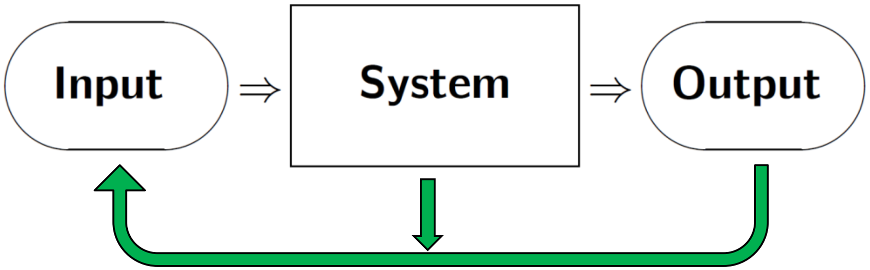
Scientific Computing & Just Enough Theory

# What is an Inverse Problem?

In a forward problem, we use a mathematical model to compute the output from a "system" given the input.



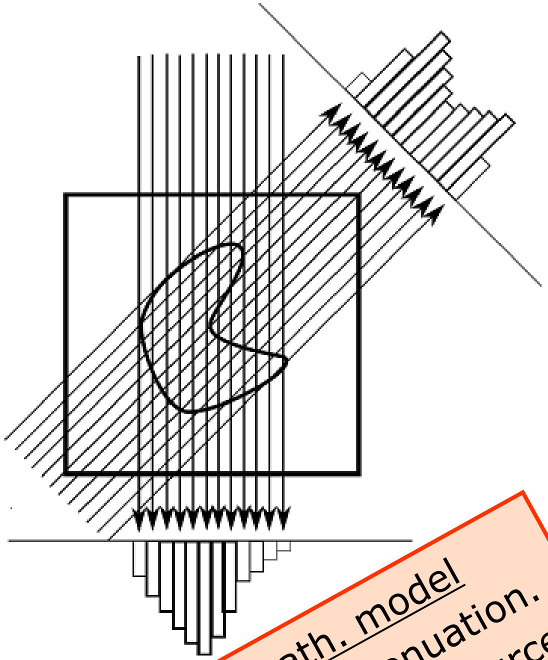
In an inverse problem we estimate a quantity that is not directly observable, using indirect measurements and the forward model.



Some examples on the next pages.

# Example: Tomography

Image reconstruction from projections.

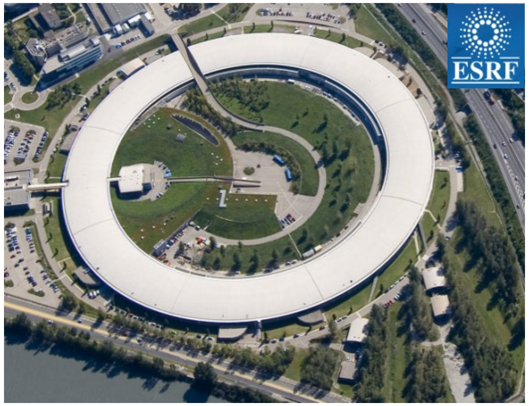


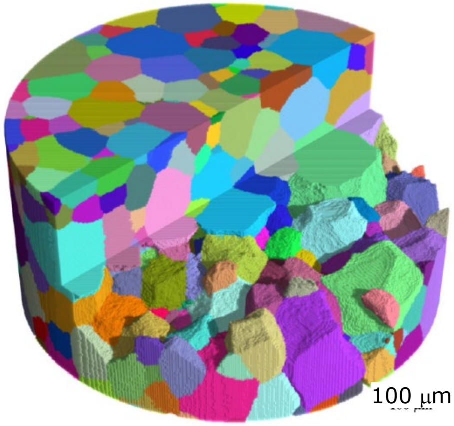
Required in the math. model

- Physics of X-ray attenuation.
- Strength of the X-ray source.
- Specification of the geometry.

Medical imaging

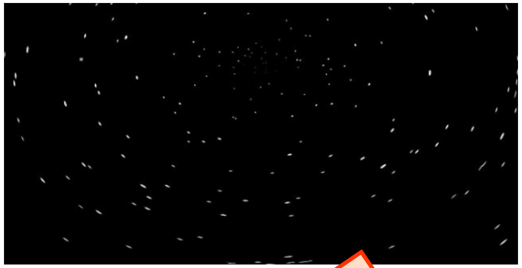
Materials science

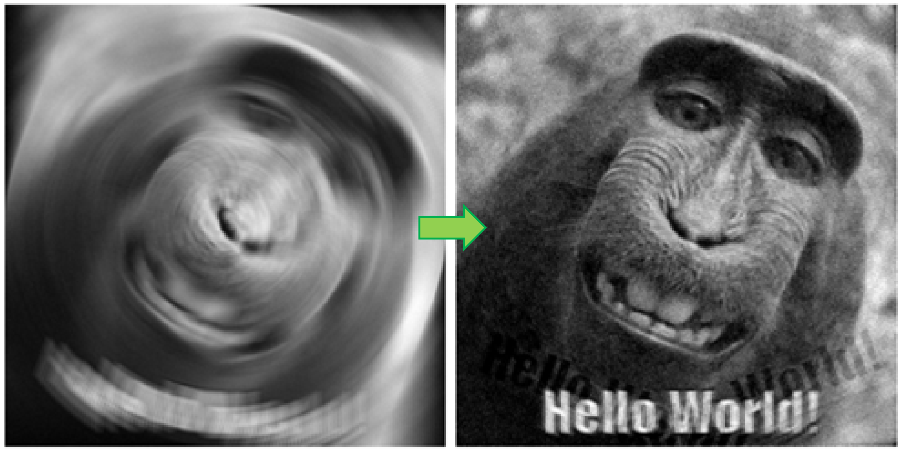




# Example: Rotational Image Deblurring

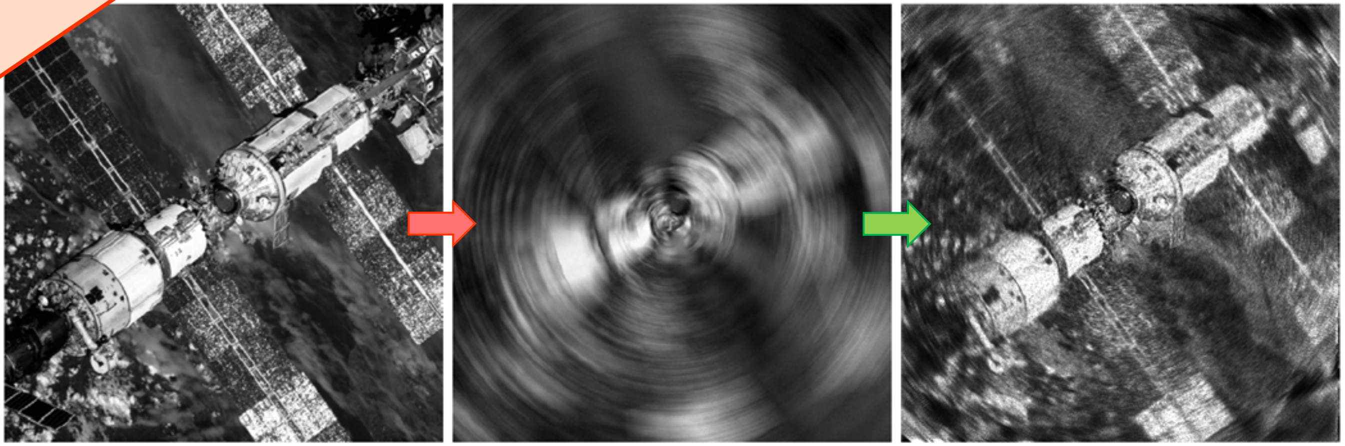
Application: “star camera” used in satellite navigation.





<u>Required in the math. model</u>

- Center of rotation.
- Rotation angle.



# Inverse Problem and Mathematics

## Inverse problems

- arise when we use a mathematical model $\mathcal{K}$
- to infer about *internal* or *hidden features* $f$
- from *external* and/or *indirect measurements* $g$.

$$\boxed{\mathcal{K} f = g}$$

## Why mathematics is important

- A solid foundation for formulation of inverse problems.
- A framework for developing computational algorithms.
- A "language" for defining and expressing the properties of the solutions: existence, uniqueness, stability, reliability, …

# Some Formulations

Mathematical formulations of inverse problems take different forms.

Fredholm integral equation of the first kind:

This talk

$$\int_0^1 K(s,t)\, f(t)\, dt = g(s) \ , \quad 0 \leq s \leq 1 \ .$$

Calderón problem (PDE with Dirichlet BC):

$$\begin{cases} \nabla \cdot \sigma\, \nabla u = 0 & u \in \Omega \\ u = f & u \in \partial\Omega \end{cases}$$

Fascinating: very different applications of inverse problems lead to the same formulations.

# A Few Simple Examples

$$\int K(s,t)\, f(t)\, dt = \frac{1}{6}\,(s^3 - s) \ , \qquad K(s,t) = \begin{cases} s(t-1) \ , & s < t \\ t(s-1) \ , & s \geq t \ . \end{cases}$$

The solution is the second derivative, so $f(t) = t$.

$$\int_0^{2\pi} K(s-t)\, f(t)\, dt = g(s) \ , \qquad f, g, K \text{ are } 2\pi\text{-periodic.}$$

This is deconvolution; the solution is formally given by

$$f(t) = \mathcal{F}^{-1}\big(\mathcal{F}(g) \,/\, \mathcal{F}(k)\big) \ , \qquad \mathcal{F} = \text{Fourier transform.}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x\,\cos\theta + y\,\sin\theta - s)\, f(x,y)\, dx\, dy = g(s,\theta)$$

$$s \in [0,1] \ , \quad \theta \in [0, 2\pi) \ .$$

This is the Radon transform underlying X-ray CT.

# Eigenvalue Analysis for Symmetric Kernel

$$\boxed{\int_0^1 K(s,t)\, f(t)\, dt = g(s) = 1 \ , \qquad 0 \leq s \leq 1 \ .}$$

A symmetric kernel $K(s,t) = K(t,s)$ has a real eigensystem,

$$\int_0^1 K(s,t)\, v_i(t)\, dt = \lambda_i\, v_i(s) \ , \qquad i = 1, 2, 3, \ldots$$

Then we can write the solution as

$$f(t) = \sum_{i=1}^{\infty} \frac{\langle v_i, g \rangle}{\lambda_i}\, v_i(t) \ , \qquad \langle v_i, g \rangle = \int_0^1 v_i(s)\, g(s)\, ds \ .$$

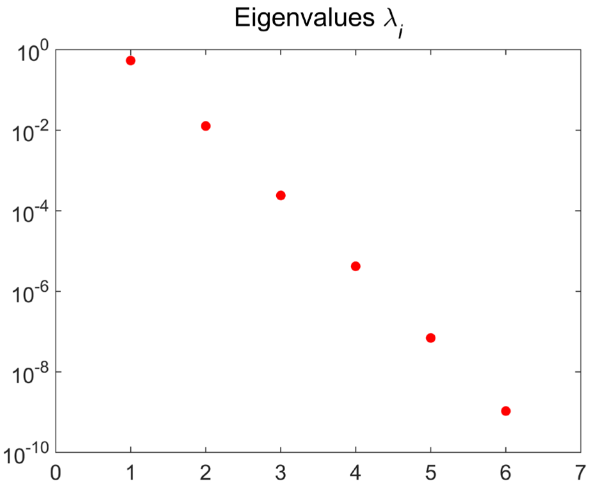
This is very useful for analysis of inverse problems (but no so much for numerical computations).

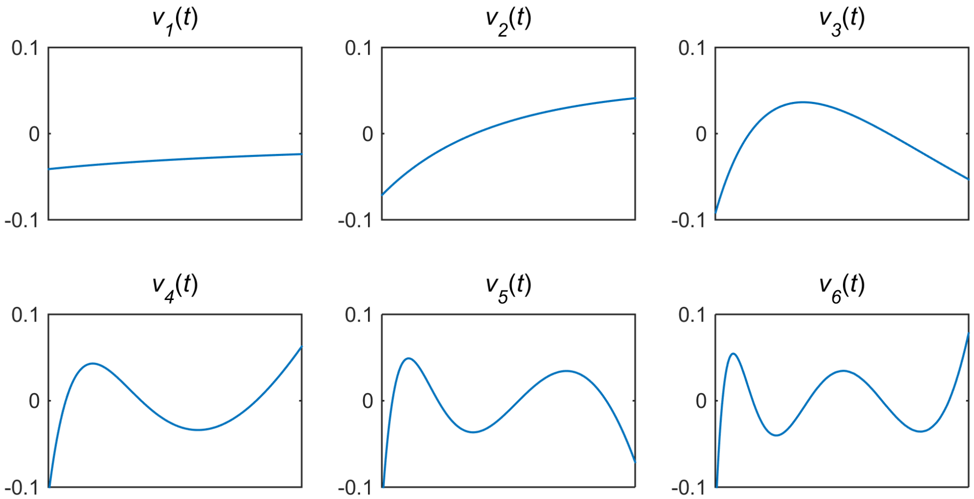
# A Tricky Example …

$$\boxed{\int_0^1 \frac{1}{s+t+1}\, f(t)\, dt = g(s) = 1 \ , \qquad 0 \le s \le 1 \ .}$$

Can you guess a solution?

Eigenvalues and eigenfunctions:

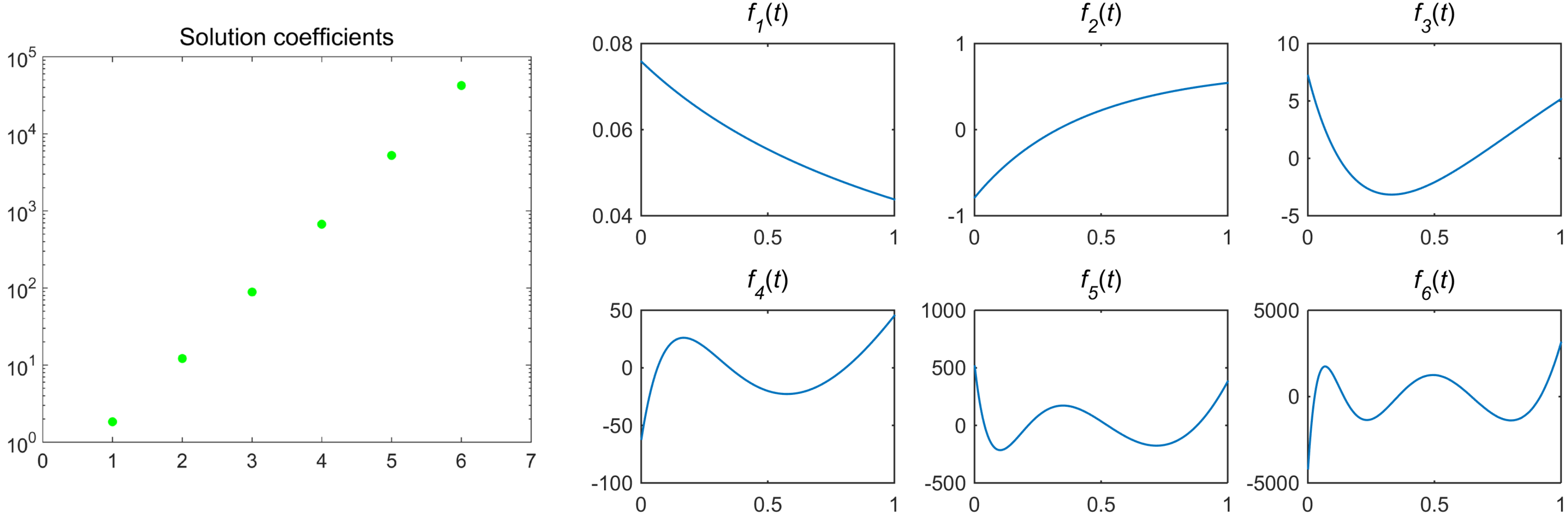




# … With No Solution

Let us compute finite approximations to the solution:

$$f_k(t) \equiv \sum_{i=1}^{k} \frac{\langle v_i, g \rangle}{\lambda_i} \, v_i(t) \; , \qquad i = 1, 2, 3, \ldots$$



The amplitude of $f_k(t)$ becomes disturbingly large as $k$ increases, and the sum does not converge as $k \to \infty$.

# Inverse Problems Are Ill Posed

Hadamard's definition of a well-posed problem (early 20th century)

1. Existence: the problem must have a solution.
2. Uniquness: the solution must be unique.
3. Stability: it must depend continuously on data and parameters.

If the problem violates any of these requirements, it is ill posed.

Inverse problems are, by nature, always ill posed.

And yet, we have a strong desire – and a need – to solve them …

# Hadamard 1 (existence) and 2 (uniqueness)

**Case 1**

$$A\,x = b \quad \Leftrightarrow \quad \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2.1 \\ 3.0 \\ 3.9 \end{pmatrix}$$

There is no $x$ that satisfies this equation, but we can define the least squares solution that minimizes the residual norm

$$x_{\mathsf{LS}} \equiv \mathrm{argmin}_x \|A\,x - b\|_2 = \begin{pmatrix} 1.2 \\ 0.9 \end{pmatrix}$$

**Case 2**

$$A\,x = b \quad \Leftrightarrow \quad \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} .$$

There are infinitely many $x$ that satisfy this equation; we can define the unique minimum-norm solution that minimizes the solution's norm

$$x_0 \equiv \mathrm{argmin}_x \|x\|_2 \quad \text{s.t.} \quad A\,x = b \qquad \Rightarrow \qquad x_0 = \begin{pmatrix} 0.6 \\ 1.2 \end{pmatrix} .$$

# Hadamard 3 (stability)

Unperturbed system:

$$A = \begin{pmatrix} 1.0 & 2.1 & 3.0 \\ 4.0 & 5.0 & 5.9 \\ 7.0 & 8.0 & 9.0 \end{pmatrix}, \quad x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad b = A\,x = \begin{pmatrix} 6.1 \\ 14.9 \\ 24.0 \end{pmatrix}.$$

Perturbed system:

$$\tilde{b} = b + \begin{pmatrix} 0 \\ 0.001 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \tilde{x} = A^{-1}\tilde{b} = \begin{pmatrix} 0.927 \\ 1.171 \\ 0.904 \end{pmatrix}.$$

The matrix $A$ is ill conditioned, $\mathrm{cond}(A) = 4249$, and therefore the solution is very sensitive to perturbations of $b$ and $A$:

$$\frac{\|\Delta x\|}{\|x\|} \leq \mathrm{cond}(A) \left( \frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \right).$$

# Eigenvalue Analysis for Symmetric Kernel

Recall that we can write the solution as

$$f(t) = \sum_{i=1}^{\infty} \frac{\langle v_i, g \rangle}{\lambda_i} v_i(t) \ , \qquad \langle v_i, g \rangle = \int_0^1 v_i(s)\, g(s)\, ds \ .$$

**Picard condition** for a square integrable solution:

$$\| f \|_2^2 = \sum_{i=1}^{\infty} \left( \frac{\langle v_i, g \rangle}{\lambda_i} \right)^2 < \infty \ .$$

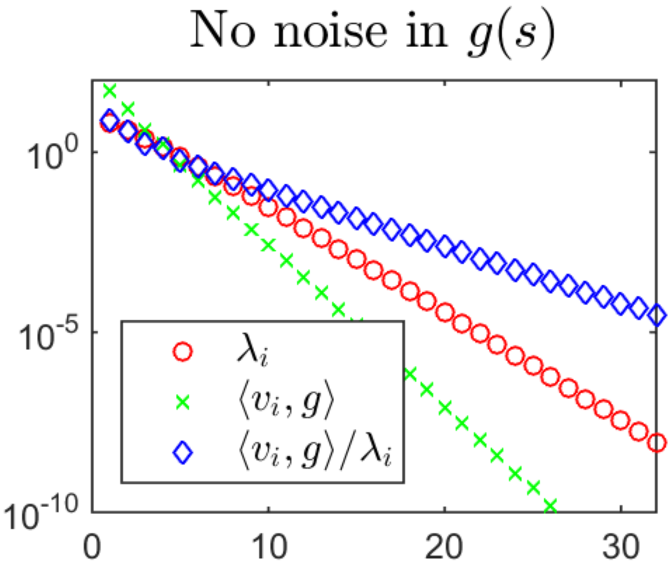
The enumerator must decay *sufficiently faster* than the denomiator.

Specifically, the coefficients $\langle v_i, g \rangle$ must decay faster than $\lambda_i \, i^{-1/2}$.

# Eigenvalue Analysis for Symmetric Kernel

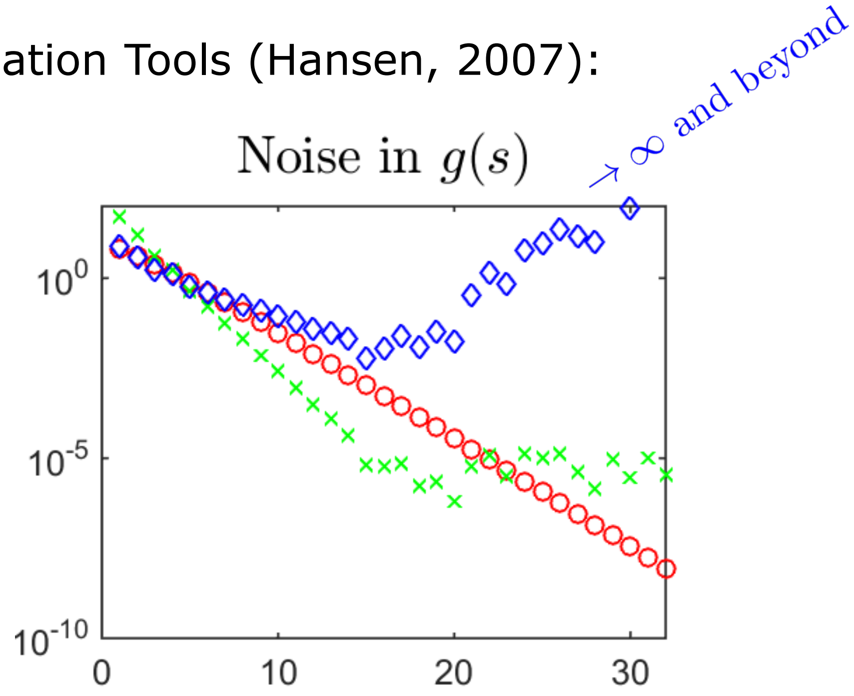
Recall: the coefficients $\langle v_i, g \rangle$ must decay faster than $\lambda_i \, i^{-1/2}$.

Test problem – **`gravity`** from Regularization Tools (Hansen, 2007):



No noise in $g(s)$

With no noise in the data, the Picard condition is satisfied.



Noise in $g(s)$

$\to \infty$ and beyond

When noise is present, the Picard condition is not satisfied.

The solution coefficients diverge.

# Dealing with the Instability → Regularization

The ill conditioning of the problem makes it impossible to compute a "naive" solution to the inverse problem:

$$K\,f = g \qquad \rightarrow \qquad \xcancel{f_{\text{naive}} = K^{-1} g}$$

→ Incorporate prior information about the solution via **regularization**:

$$\boxed{\min_f \left\{ \|K\,f - g\|_2^2 + \alpha\, R(f) \right\} \ , \qquad R(f) = \text{regularizer}}$$

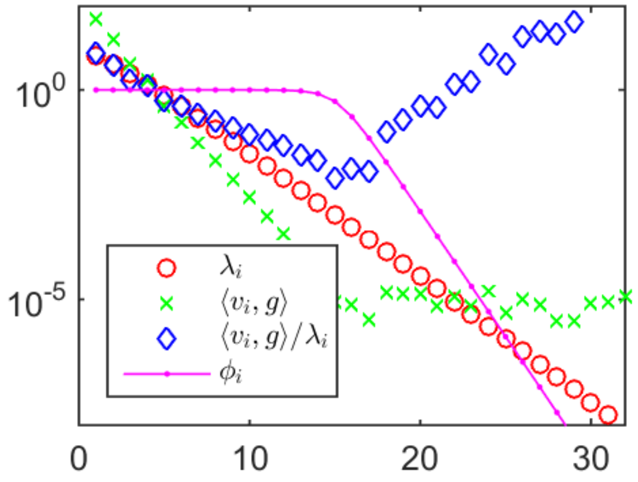
# Eigenvalue Analysis of Tikhonov Regularizer

The important special case of Tikhonov regularization

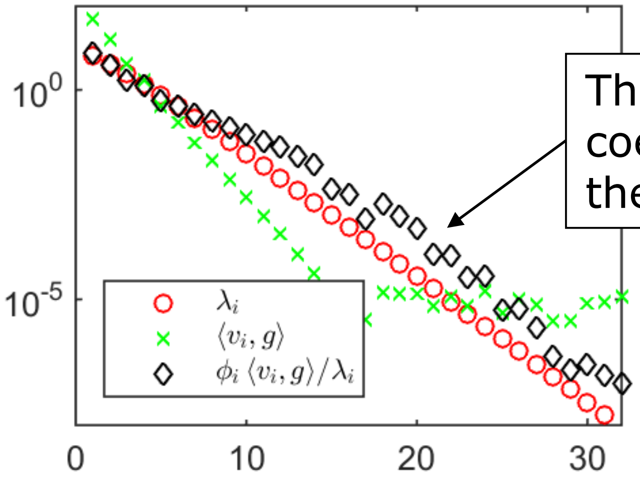
$$R(f) = \|f\|_2^2 \qquad \Rightarrow \qquad f(t) = \sum_{i=1}^{\infty} \frac{\lambda^2}{\lambda^2 + \alpha} \frac{\langle v_i, g \rangle}{\lambda_i} v_i(t) \ .$$

Here $\phi_i = \dfrac{\lambda^2}{\lambda^2 + \alpha}$ are the *filter factors*.

Noise in $g(s)$



Tikhonov



These *modified* coefficients satisfy the Picard condition.

Stabilization accomplished!

# Case: Total Variation (TV)

Prior: image consists of regions with constant intensity and sharp edges. How to say this *in mathematical terms?*

Total variation regularization term $R(f) = \int_\Omega \|\nabla f\|_2 \, d\Omega \rightarrow$

$$\boxed{R(x) = \textstyle\sum_{\text{pixels}} \|D_i \, x\|_2 \ , \quad D_i \, x = \text{gradient}}$$

# Case: Directional TV (DTV)

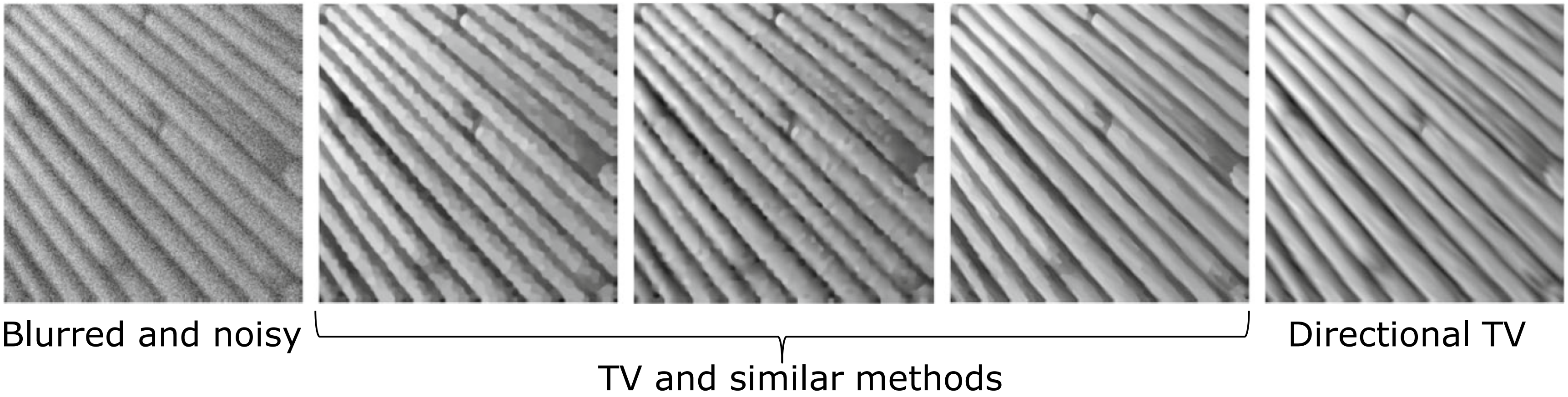
Kongskov, Dong, Knudsen, *Directional total generalized variation regularization*, 2019.

Prior: the edges in the mage have a dominating direction $\theta$.
How to say that in mathematical terms?

Directional TV regularization term:

$$\boxed{R(f) = \int_\Omega \left\| \begin{pmatrix} D_\theta f \\ \gamma D_{\theta^\perp} f \end{pmatrix} \right\|_2 d\Omega \ ,}$$

where $D_\theta$ is the directional derivative.



# Case: Regularization with Sparsity Prior

TV = a ”sparsity prior” that produces a solution with a *sparse gradient*.

We can also require that the *solution itself is sparse*, i.e., the image has many nonzero pixels. How to say this in mathematical terms?

Use the 1-norm to enforce sparsity:

Candès, Donoho, Romberg, Tao.

$$\boxed{R(x) = \|x\|_1 = \textstyle\sum_i |x_i| \ .}$$

This is well known from **compressed sensing** where it is succesfully used to reconstruct a sparse signal $x$ from limited data.

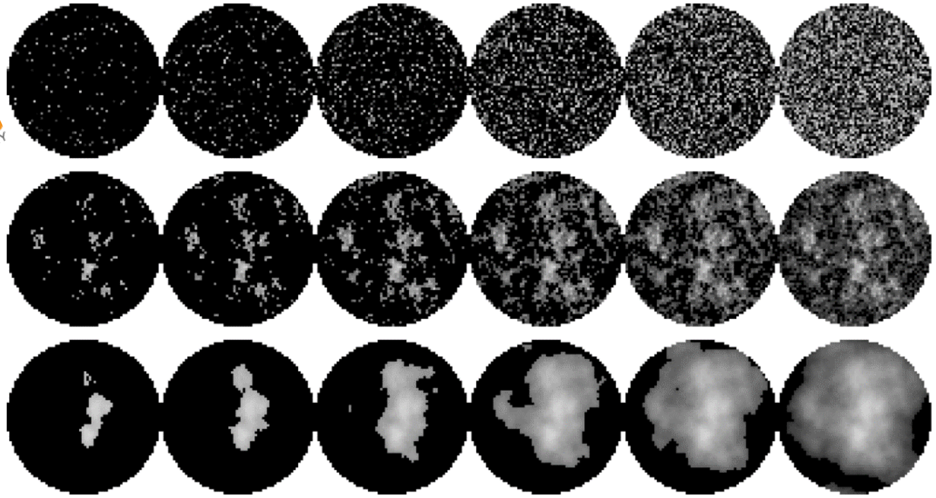
THEOREM: we can reconstruct a sparse $x \in \mathbb{R}^n$ with at most $p$ nonzeroes from a data vector $b \in \mathbb{R}^m$ with $b = A\,x$ if $A$ is random and $m \approx 2p$.

In our inverse problems, $A$ is certainly not random – it is a discretization of the forward operator.

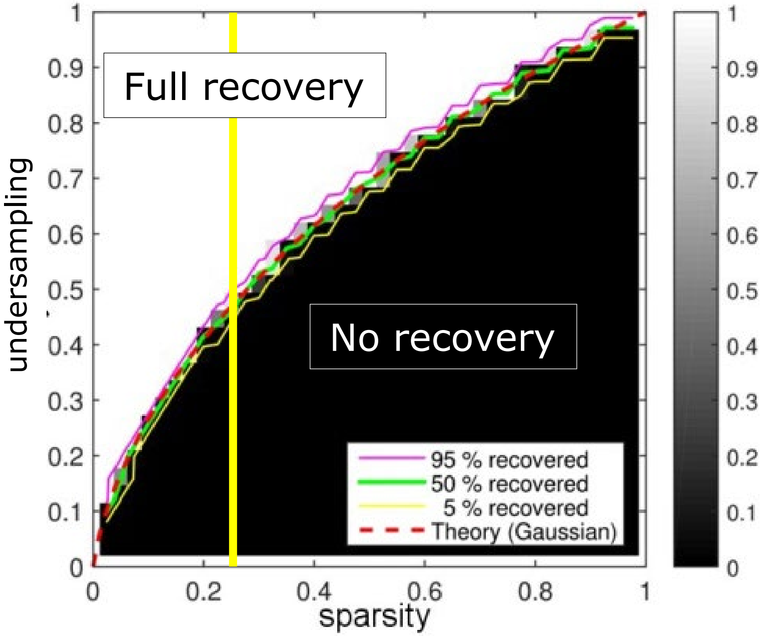
Surprisingly, we can still use a 1-norm regularization term $R(x) = \|x\|_1$.

# Case: Sparse CT Reconstruction

Jørgensen, Sidky, H, Pan, *Empirical Average-Case Relation Between Under-sampling and Sparsity in X-Ray CT*, 2015.



Artificial sparse test images.
Left to right: 5%, 10%, 20%, 40%, 60%, 80% nonzeroes.



**Phase diagram:** the *recovery fraction* of reconstructed images at a given *sparsity* abruptly changes from 0 to 1, once a critical number of measurements is reached.

Agrees with the theoretical phase transition for random matrices (Donoho, Tanner 2009).

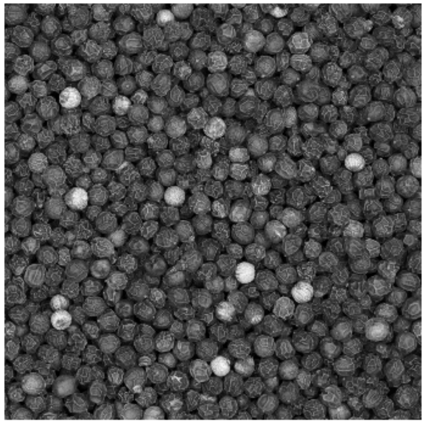
# Case: Training Images as Regularizer

Soltani, Kilmer, H, *A tensor-based dictionary learning approach to tomographic image reconstruction*, 2016.

Soltani, Andersen, H, *Tomographic image reconstruction using training images*, 2017.

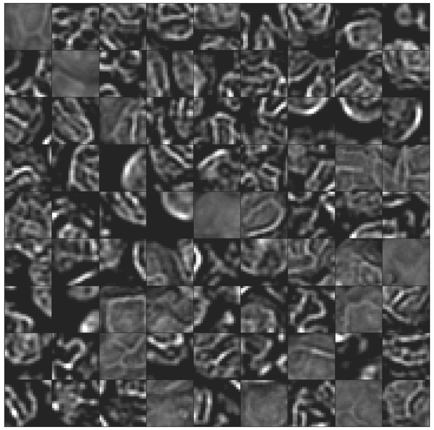
Training images

are *patches* from high-res image.



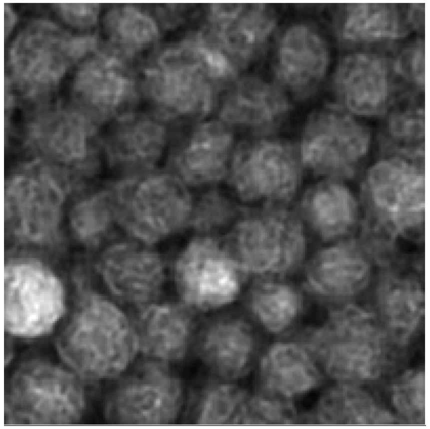
Dictionary patches

*learned* via nonneg. matrix factorization.



Reconstruction

computed from highly *underdet*. problem.



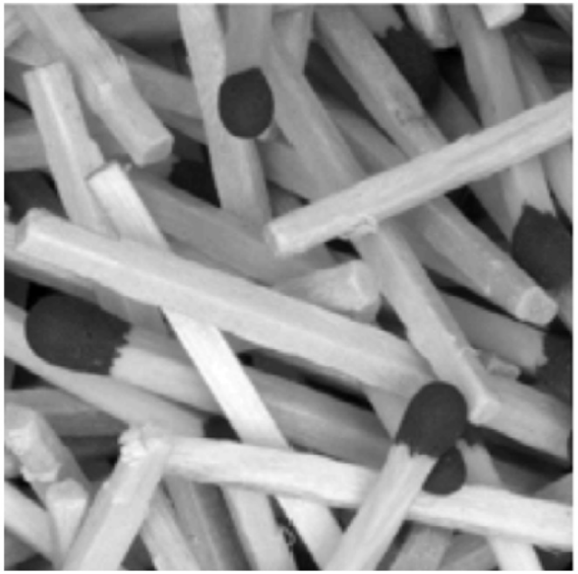
Dictionary

Sparsity prior on dictionary elements

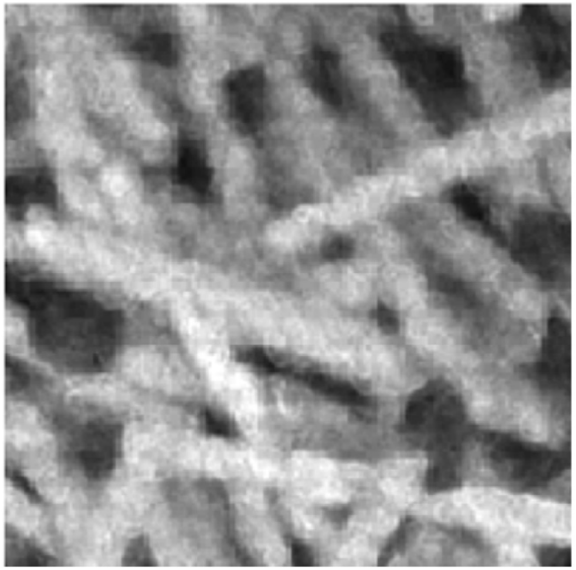
$$\min_z \| A\,W z - b\|_2^2 + \alpha \|z\|_1 \ , \qquad x = W z \ .$$

# Case: When the Training Images are Wrong

Soltani, Andersen, H, *Tomographic image reconstruction using training images*, 2017.



Exact image



The "best" reconstruction based on a wrong dictionary created from the peppers training image.

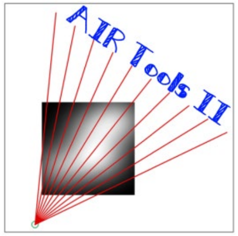
Peppermatches?

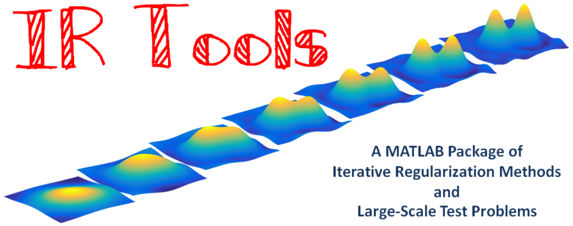
# Algorithm Development – Iterative Methods

*Large-scale problems $A\,x = b$.*

How to solve them efficiently?

→ Iterative methods!





Gradient (steepest descent) method for computing CT solutions:

$$\boxed{x^k \leftarrow x^{k-1} + \omega\, B(b - A\, x^{k-1})\ .}$$

Here $A$ = Radon transform = forward projector, and $B$ = backprojector.

By definition, $B = A^T$ and $x^k$ converges to the least squares solution.

So who in their right mind would write software where $B \neq A^T$?

All good HPC-programmers! Efficient use of GPUs etc.

*Need to study the implications of this fact.*

# Convergence Explained

Convergence of an iterative method means that the sequence of iteration vectors

$$x^0 \to x^1 \to x^2 \to x^3 \to \cdots$$

approaches a limit vector as $k \to \infty$.

For the steepest descent method (with $B = A^T$) we have

$$x^k \to x_{\mathsf{LS}} \qquad \text{for} \qquad k \to \infty \ ,$$

where $x_{\mathsf{LS}}$ is the least squares solution.

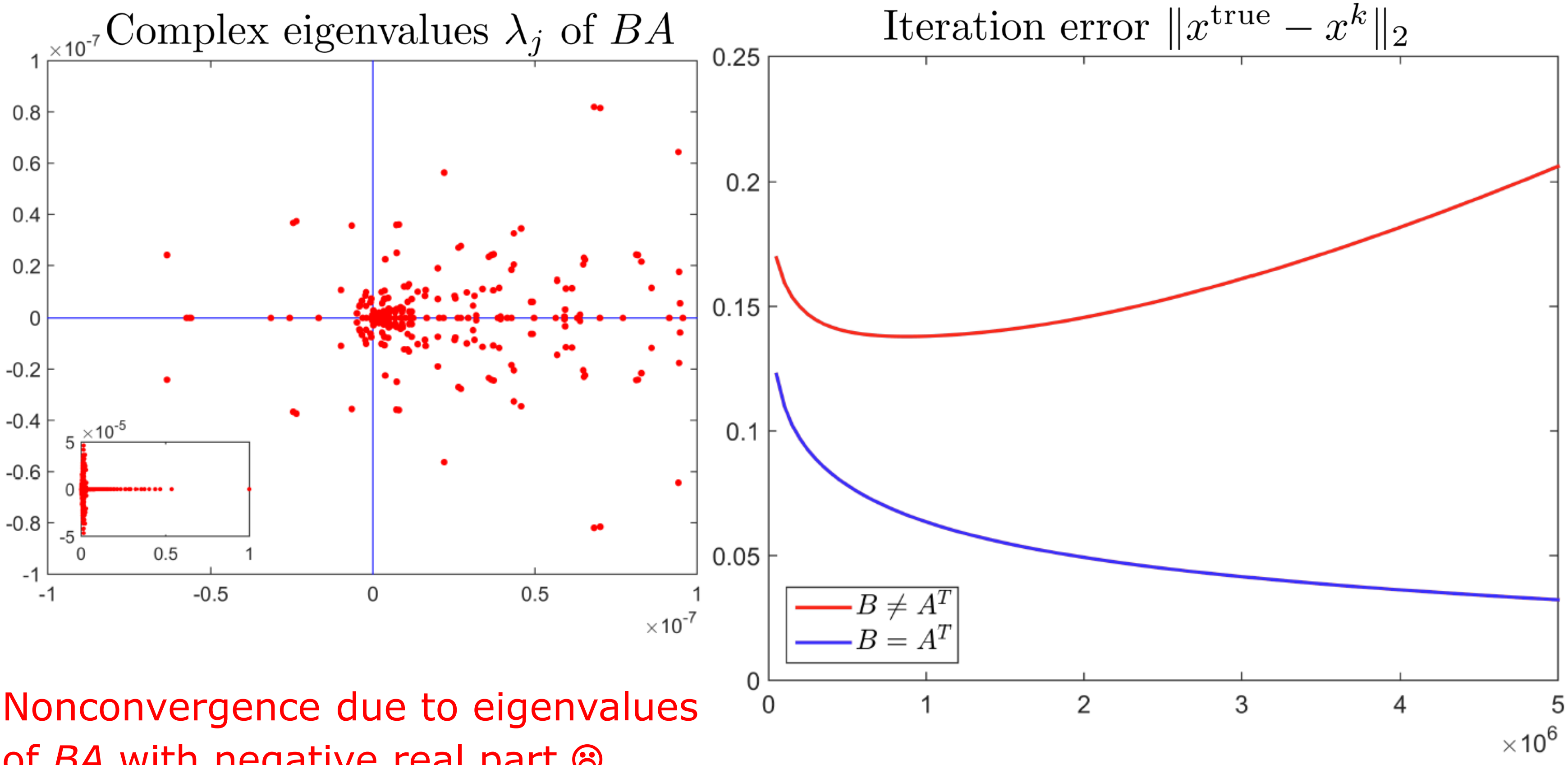
The *conditions* for convergence are:

$$0 < \omega < 2\, \frac{\mathrm{Re}\, \lambda_j}{|\lambda_j|^2} \qquad \text{and} \qquad \mathrm{Re}\, \lambda_j > 0 \ ,$$

where $\lambda_j$ are the eigenvalues of $BA$.

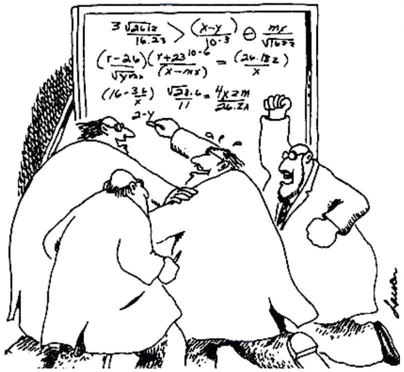
# Nonconvergence!

Parallel-beam CT, unmatched pair from *ASTRA*, 64 × 64 Shepp-Logan phantom, 90 projection angles, 60 detector pixels, $\min \mathrm{Re}\, \lambda_j = -6.4 \cdot 10^{-8}$.



Nonconvergence due to eigenvalues of *BA* with negative real part ☹

# The Fix



1. Ask the software developers to change their implementation of $B$ and/or $A$?
   $\rightarrow$ Significant loss of comput. efficiency.

2. Use mathematics to *fix* the nonconvergence.

We define the **shifted** version of the iterative algorithm:

$$x^{k+1} = (1 - \sigma\,\omega)\, x^k + \omega\, B\, (b - A\, x^k) \ , \qquad \alpha > 0$$

with just one extra factor $(1 - \sigma\,\omega)$; simple to implement.

Condition for convergence:

$$0 < \omega < 2\, \frac{\mathrm{Re}\,\lambda_j + \sigma}{|\lambda_j|^2 + \sigma\,(\sigma + 2\,\mathrm{Re}\,\lambda_j)} \qquad \text{and}$$

$\mathrm{Re}\,\lambda_j + \sigma > 0$ .

Choose the shift $\sigma$ *just large enough!*

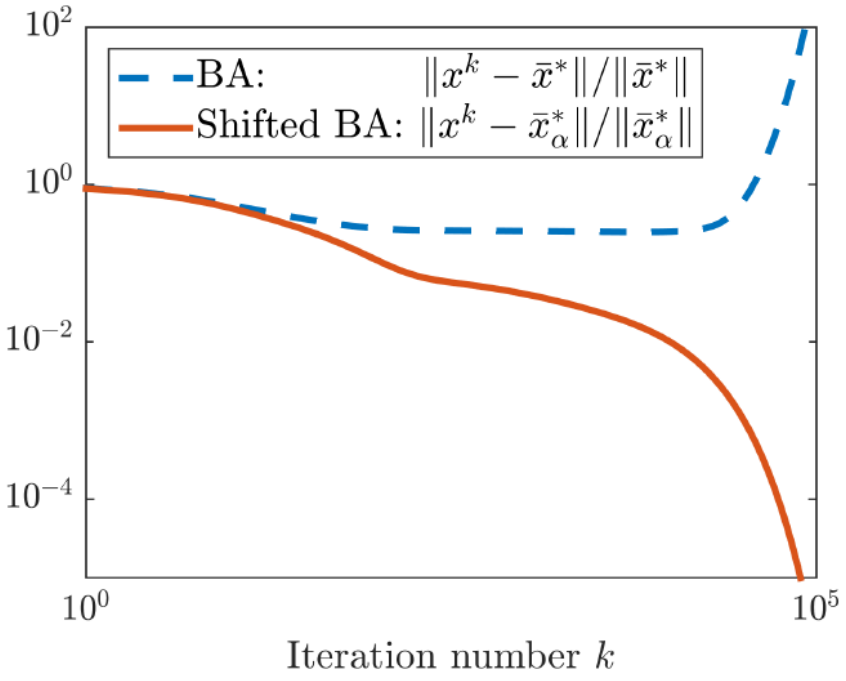
Dong, H, Hochstenbach, Riis; SISC, 2019.

I am really proud of this paper.

# Nonconvergende → Convergence

Parallel-beam CT, 90 projections in the range 0°–180°, 80 detector pixels; $128 \times 128$ Shepp-Logan phantom; $m = 7\,200$ and $n = 16\,384$.
Both $A$ and $B$ are generated with the GPU-version of the ASTRA toolbox.

$\rho(BA) = 1.76 \cdot 10^4$
$\alpha = 1.85$

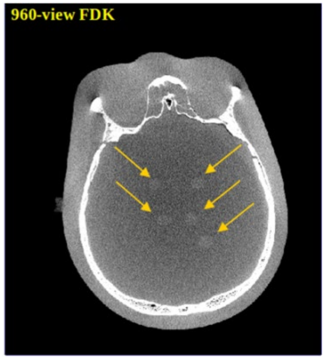


The BA Iteration diverges from $\bar{x}^* = (BA)^{-1}Bb$.

The Shifted BA Iteration converges to fixed point $\bar{x}^*_\alpha = (BA + \alpha\, I)^{-1}Bb$.

# Beyond Sharp Reconstructions → CUQI

Classical method.



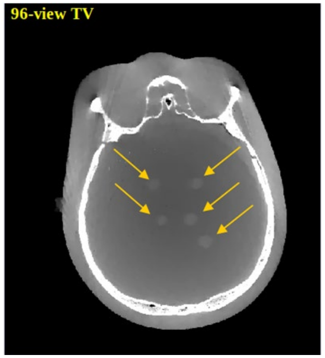


*Figure credit to E. Sidky*

TV regularization needs only 10% of full X-ray dose.

But how reliable are the spots?

All kinds of errors have influence on the solution:

$$x = \text{argmin} \left\{ ||\mathcal{K} f - g|| + \text{regularization}(x) \right\}$$

algorithm-error

model-error

data-error

regularization-error

**UQ = uncertainty quantification** is the end-to-end study of the impact of all forms of error and uncertainty in the data and models.

Computational Uncertainty Quantification for Inverse Problems

A research initiative sponsored by Villum Fonden (the Villum Foundation)

CUQI

Picture taken by Per Christian Hansen in the garden of the Heian Shrine in Kyoto, Japan.

# Applied UQ

Black hole?

Traditionally: one result.

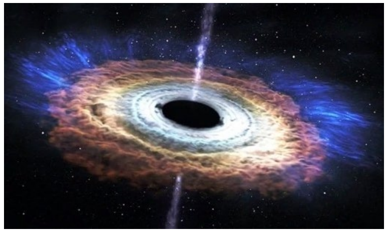
How trustworthy is it?

UQ gives insight about the reliability of the result.

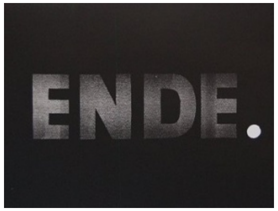
**12%**

**17%**

**63%**



**8%**



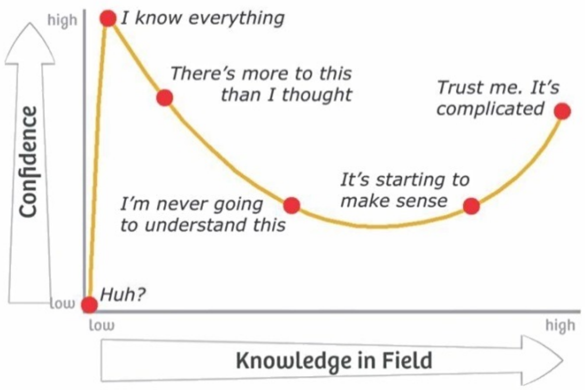
# Research Initiative CUQI

## Computational Uncertainty Quantification for Inverse Problems

- Develop the mathematical, statistical and computational framework.
- Create a modeling framework and a computational platform for non-experts.

**Vision**

Computational UQ becomes an essential part of solving inverse problems in science and engineering.



# UQ: Gaussian Data Errors and Gaussian Prior

Model: $b = A\,\bar{x} + e$ with $A \in \mathbb{R}^{m\times n}$ fixed and $e = \mathcal{N}(0,\sigma^2 I)$.

The pdf for $b$, given $x$ and $\sigma$ (known as the *likelihood*):

$$p(b|x,\sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{m/2} \exp\left(-\frac{1}{2\sigma^2}\,\|A\,x - b\|_2^2\right) .$$

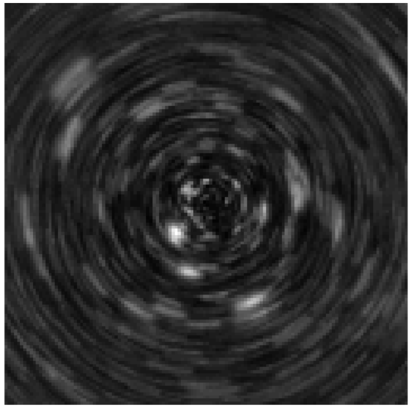
The unknown $x$ is a random vector. Assume a Gaussian prior $x \sim \mathcal{N}(0,\delta^{-1}I)$ this yields the prior

$$p(x|\delta) = \left(\frac{\delta}{2\pi}\right)^{n/2} \exp\left(-\frac{\delta}{2}\,\|x\|_2^2\right) .$$

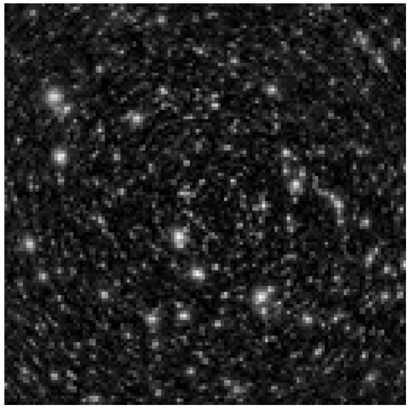
Bayes rule/law/theorem defines the *posterior* for $x$:

$$\begin{aligned} p(x|b,\sigma,\delta) &= \frac{p(b|x,\sigma)\,p(x|\delta)}{p(b|\sigma,\delta)} \propto p(b|x,\sigma)\,p(x|\delta) \\ &\propto \text{const}\cdot\exp\left(-\frac{1}{2\sigma^2}\,\|A\,x-b\|_2^2\right)\cdot\exp\left(-\frac{\delta}{2}\,\|x\|_2^2\right) \\ &\propto \exp\left(-\|A\,x-b\|_2^2 - \alpha\,\|x\|_2^2\right) , \qquad \alpha = \delta\,\sigma^2 . \end{aligned}$$

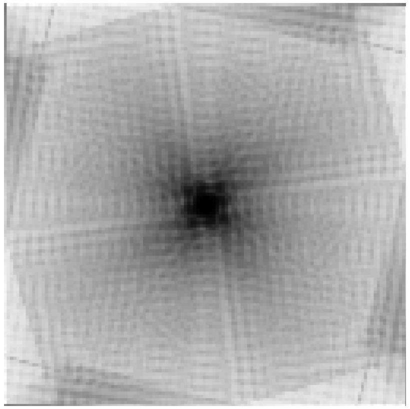
# UQ in Image Deblurring



Measured blurred image.



A solution (MAP estimator).

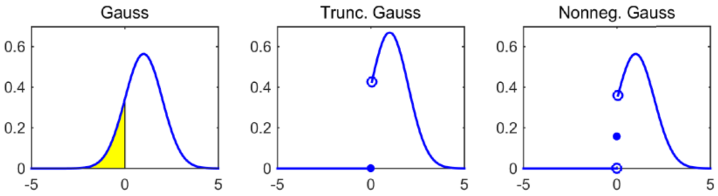


UQ shows uncertainty in each pixel; white denotes high uncertainty.

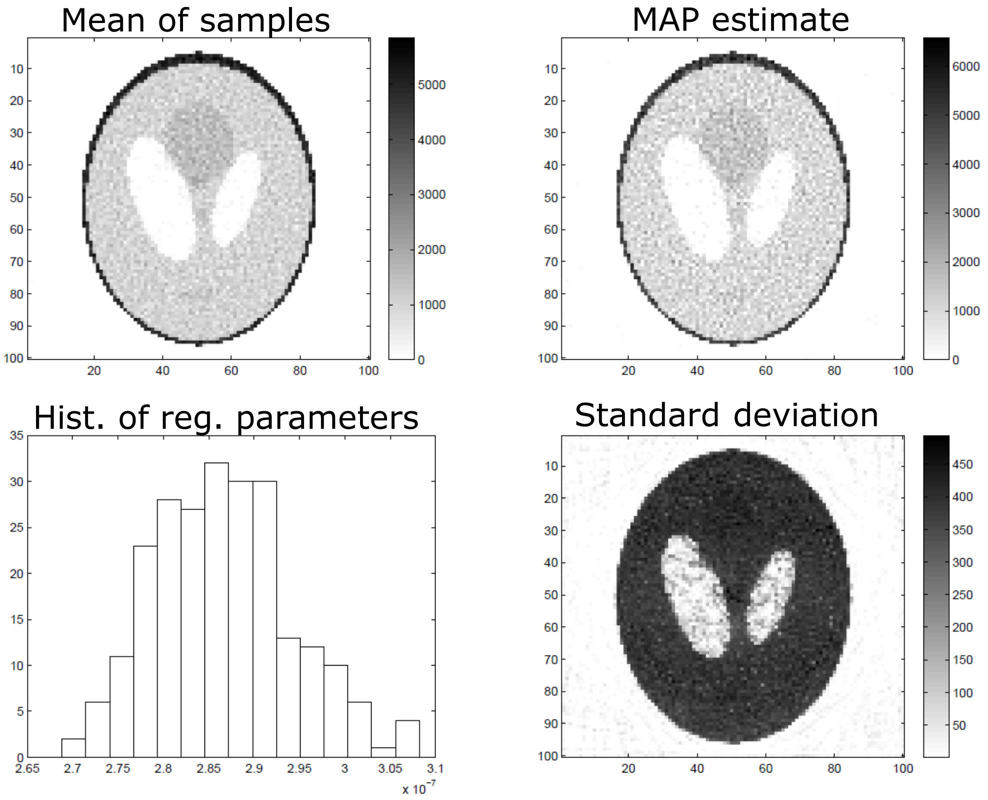
# Case: UQ with Non-Negative Prior

If the prior or likelihood is non-Gaussian, we must **sample** the posterior: we generate <u>many</u> random instances of the regularized solution with the specified likelihood and prior.

Bardsley, Hansen, *MCMC Algorithms for Non-negativity Constrained Inverse Problems*, 2019.


Gauss
Trunc. Gauss
Nonneg. Gauss

We have an analytical expression for the prior, but no analytical expression for the posterior.

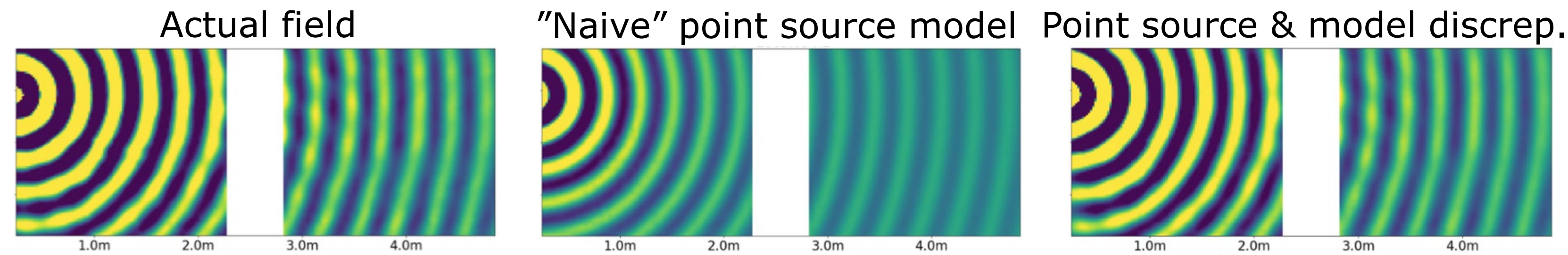

Mean of samples
MAP estimate
Hist. of reg. parameters
Standard deviation

*Positron Emission Tomography.*

Solutions sampled by a new Poisson Hierarchical Gibbs Sampler.

# Case: UQ for Model Discrepancies

Dong, Riis, Hansen, *Modeling of sound fields*, joint with DTU Elektro, 2019.



Actual field

"Naive" point source model

Point source & model discrep.

Described by a Gaussian process

Measured data = Physical model + Model discrepancy + Data errors

Physical model: Cannot include all possible aspects

Model discrepancy: Accounts for *known unknowns* & *unknown unknowns*

Data errors: Known statistics

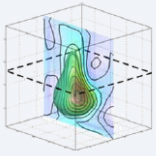
# HD-Tomo: High-Definition Tomography

The following examples are from the project **HD-Tomo**, which was funded by an ERC Advanced Research Grant, 2012–17.

## Objective: Optimal Use Prior Information

*Tomographic imaging* allows us to see inside objects. Doctors look for cancer, physicists study microscopic details of materials, security personnel inspect luggage, engineers identify defects in pipes, concrete, etc.

To achieve **high-definition tomography**, sharp images with reliable details, we must use *prior information* = *accumulated knowledge about the object.* **This project: how to do this in an optimal way.**



## Outcome: Insight, Framework and Algorithms

We developed *new theory* that provides insight and understanding of the challenges and possibilities of using advanced priors.

This insight allowed us to develop a *framework for precisely formulated tomographic algorithms* that produce *well-defined results*.

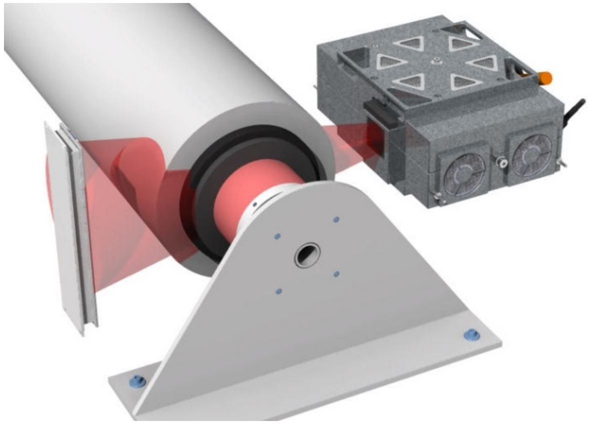
We laid the groundwork for the next generation of algorithms that will further optimize the use of prior information.

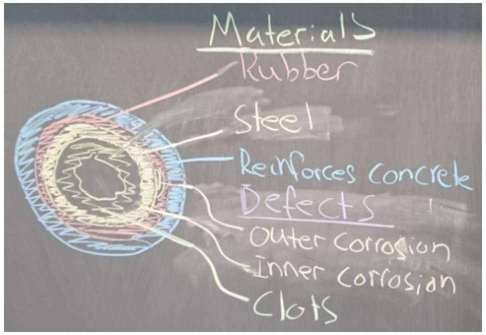
The project produced **47** journal papers, **6** proceeding papers, **7** software packages, **25** bachelor/master projects and **3** workshops.

# Example: Fault Inspection

Use X-ray scanning to compute cross-sectional images of oil pipes on the seabed.

Detect *defects*, *cracks*, etc. in the pipe.





<u>Required in the math. model</u>

- Strength of the X-ray source.
- Specification of the geometry.
- Structure of the oil pipe.