Finnish Centre of Excellence
in Inverse Problems Research

Inverse Problems in the Bayesian Framework

Erkki Somersalo

Helsinki University of Technology

Inverse problems and statistical inference

J. Kaipio and E. Somersalo: *Statistical and Computational Inverse Problems*, Springer 2004.

- Inverse problems: Concerns the problem of retrieving information of unknown parameters by indirect observations.
- Statistical inference: Concerns the problem of inferring properties of an unknown distribution from data generated from that distribution.

Why Statistics?

"Statistics is the science of information gathering, especially when the information arrives in little pieces rather than in one or two big pieces."

(Bradley Efron)

Bayesian perspective to inverse problems

- All unknowns are modelled as random variables.
- Randomness is an expression of the *lack of information*, or *ignorance* of their values.
- Random variables are characterized by their probability distributions.
- Inverse problem: Find the probability distribution of the unknowns you are interested in.

Note 1: Randomness is not the object's but the *subject's* property[1].

Note 2: Computational models predict only observables, i.e., what an observer (subject) can expect.[2]

[1] Bruno de Finetti: "Probability does not exist!"

[2] Niels Bohr: "It is a mistake to think that physics should reveal how the nature is made. Physics deals with what can be said about the nature."

Introductory example

Classical linear inverse problem: estimate $x \in \mathbb{R}^n$ from the measurement of

$$y = Ax + \text{noise}, \quad y \in \mathbb{R}^m, \quad A \in \mathbb{R}^{m \times n}.$$

Observations:

- Evidently, the value of the noise is not known, so it is natural to model it as a random variable. (All usually agree about this).
- Hence, the data y must be a realization of a random variable.
- What was the motivation for modelling the noise as random variable?

  That we do not know its value!

  Do we know the value of x?

  No!

  So, it equally natural to model x as random variable, too.

Stochastic extension of the model

Write

$$Y = AX + E,$$

where X, Y and E are random variables.

$$\begin{aligned} \text{Data} &= y_{\text{measured}} \\ &= \text{realization of } Y. \end{aligned}$$

Fundamental question: What is the probability distribution of X when the measured value for Y has realized?

Probability distributions and densities

Let X be a random variable with values in $\mathbb{R}^n$.

$$\begin{aligned} \mathrm{P}\{X \in B\} &= \text{probability that } X \in B \\ &= \mu_X(B). \end{aligned}$$

The measure μ_X is the *probability distribution of* X.

We assume that it is possible to construct a function $\pi_X(x) = \pi(x) \geq 0$ such that

$$\mu_X(B) = \int_B \pi(x)dx.$$

The function $\pi(x)$ is the *probability density of* X.

Joint probability density

Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$ be two random variables. The *joint probability density of X and Y*, denoted by $\pi(x, y)$, is defined as

$$\mathrm{P}\{X \in B_1 \text{ and } Y \in B_2\} = \int_{B_1} \int_{B_2} \pi(x, y) dx dy.$$

Two extremal cases are of particular interest:

1. When B_1 or B_2 become the full space,
2. When B_1 or B_2 shrink to one point.

Marginal densities

$X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$ random variables. Assume that Y can take *any* value:

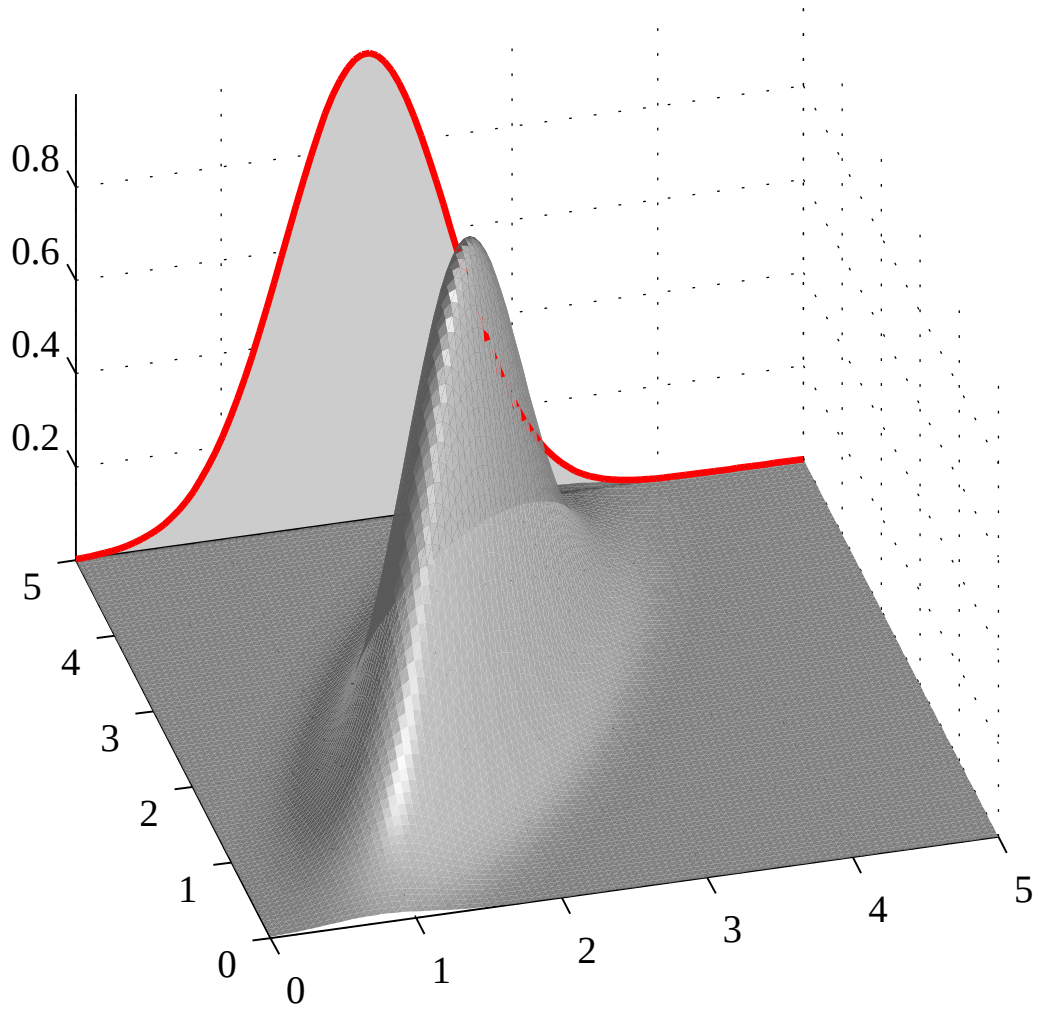
$$\begin{aligned} \mathrm{P}\{X \in B_1 \text{ and } \underbrace{Y \in \mathbb{R}^m}_{\text{always true}}\} &= \int_{B_1} \left(\int_{\mathbb{R}^m} \pi(x,y)dy \right) dx \\ = \mathrm{P}\{X \in B_1\} &= \int_{B_1} \pi(x)dx. \end{aligned}$$

Conclusion:

$$\pi(x) = \int_{\mathbb{R}^m} \pi(x,y)dy.$$

This is the *marginal density of* X.

Marginal density



Conditional density

Probability that $X \in B_1$ *provided that* $Y \in B_2$ is denoted by

$$\mathrm{P}\{X \in B_1 \mid Y \in B_2\}.$$

Since both $X \in B_1$ and $Y \in B_2$ must happen, we deduce that

$$\mathrm{P}\{X \in B_1 \mid Y \in B_2\} = \alpha\, \mathrm{P}\{X \in B_1 \text{ and } Y \in B_2\}.$$

On the other hand, if $B_1 = \mathbb{R}^n$ the probability has to be one ($X \in \mathbb{R}^n$, no matter what Y does), that is,

$$\begin{aligned} \mathrm{P}\{X \in \mathbb{R}^n \mid Y \in B_2\} &= \alpha \mathrm{P}\{\underbrace{X \in \mathbb{R}^n}_{\text{always true}} \text{ and } Y \in B_2\} \\ &= \alpha \mathrm{P}\{Y \in B_2\} = 1, \end{aligned}$$

so

$$\alpha = \frac{1}{\mathrm{P}\{Y \in B_2\}}.$$

Conclusion:

$$\mathrm{P}\{X \in B_1 \mid Y \in B_2\} = \frac{\mathrm{P}\{X \in B_1 \text{ and } Y \in B_2\}}{\mathrm{P}\{Y \in B_2\}}.$$

Let B_2 shrink to a single point y.

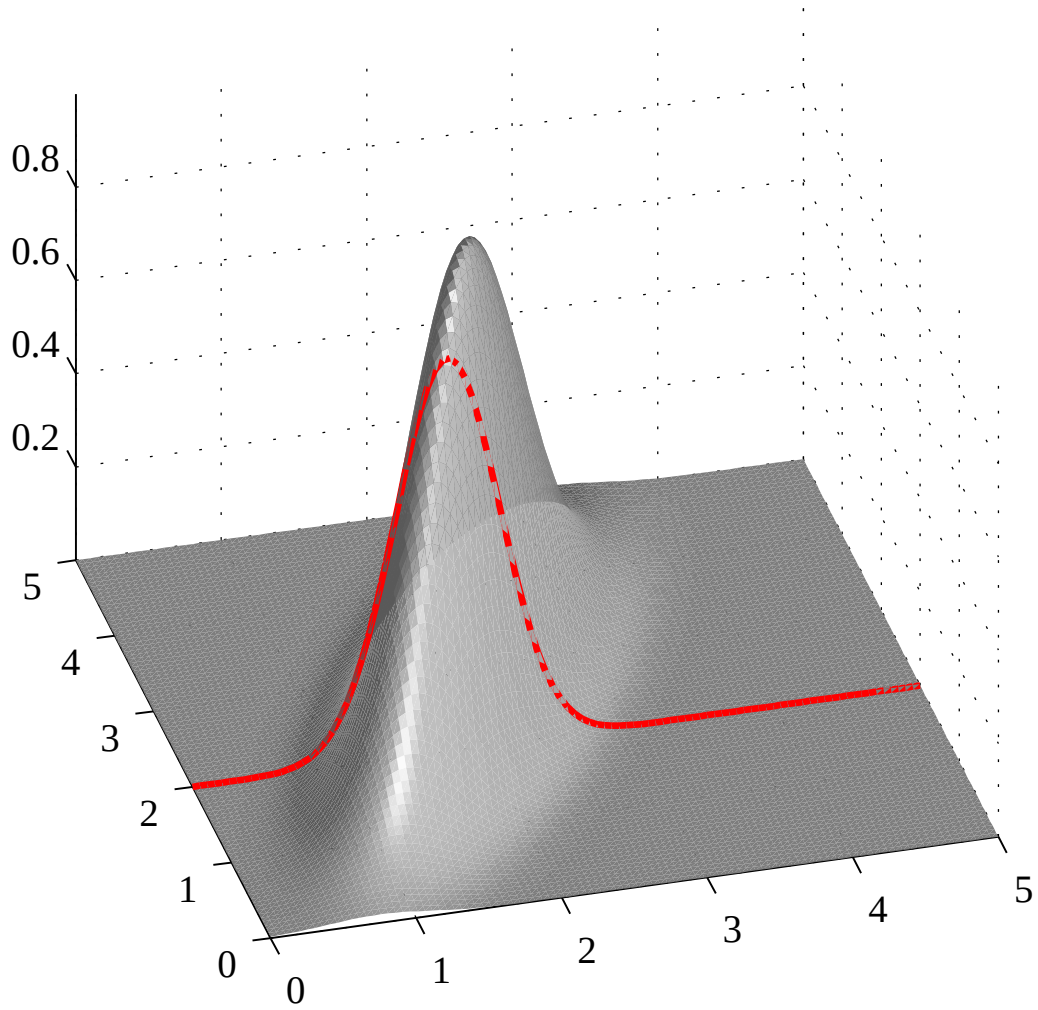
Writing the probabilities in terms of the densities, we find that

$$\mathrm{P}\{X \in B_1 \mid Y = y\} = \int_{B_1} \frac{\pi(x,y)}{\pi(y)} dx.$$

We denote the *conditional density* as

$$\pi(x \mid y) = \frac{\pi(x,y)}{\pi(y)}.$$

Conditional density



Likelihood density

Consider the additive noise model

$$Y = f(X) + E,$$

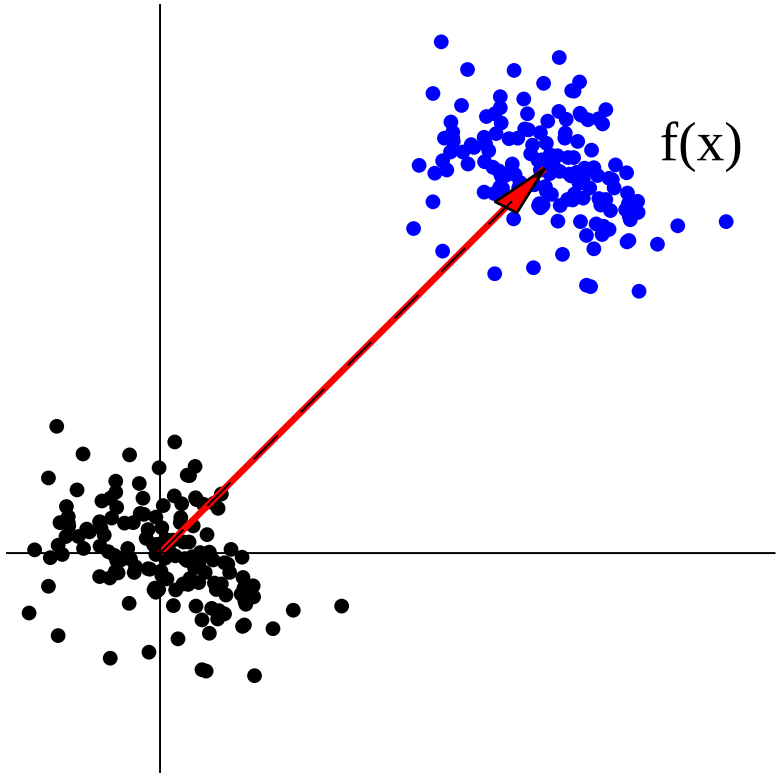
where X and E are mutually independent.

What is the probability density of Y if $X = x =$ fixed?

We have

$$Y = \underbrace{f(x)}_{\text{fixed}} + \underbrace{E}_{\text{random}},$$

so the only source of randomness in Y is the noise term.



If π_{noise} is the probability density of the noise,

$$\pi(y \mid x) = \text{likelihood function} = \pi_{\text{noise}}(y - f(x)).$$

Prior density

The prior density expresses what we believe of the values of X, whatever value the observation will later take on.

$$\pi(x) = \pi_{\text{prior}}(x) = \int_{\mathbb{R}^m} \pi(x, y)dy.$$

By the definition of the conditional density,

$$\pi(x, y) = \pi_{\text{prior}}(x)\pi(y \mid x).$$

$$\text{joint density} = \text{prior} \times \text{likelihood}.$$

Posterior density

Posterior density expresses the probability density of X when Y takes on a fixed value.

By the definition of conditional densities,

$$\pi_{\text{post}}(x) = \pi(x \mid y) = \frac{\pi(x, y)}{\pi(y)}.$$

Bayes' formula,

$$\pi_{\text{post}}(x) = \pi(x \mid y) = \frac{\pi_{\text{pr}}(x)\pi(y \mid x)}{\pi(y)}.$$

Posterior density=solution of the inverse problem.

Model problem: Imaging blocky signals and images[3,4]

Qualitative description of problem: "Given a noisy indirect observation, recover a signal or an image that **varies slowly** except for **unknown number of jumps of unknown size and location.**"

Examples: Geophysical profiling of layered Earth, optical tomography of an infant's head, atmospheric profiling of inversion layers, detection of breast cancer with EIT, localization of cortical activity by EEG, image segmentation.

Objectives:

1. Develop a fast algorithm to find the discontinuities
2. Estimate the reliability of the algorithm.

[3] D. Calvetti and E. Somersalo: Gaussian hypermodels and recover blocky objects. Inverse Problems **23** (2007) 733–754.

[4] D. Calvetti and E: Somersalo: A unified Bayesian framework for algorithms to recover blocky signals. Proc. SPIE 2007, San Diego (to appear).

Prototype problem

1D inverse problem, e.g., deconvolution, continuous model:

$$g(s_j) = \int_I \mathcal{A}(s_j, t) f(t) dt + e_j, \quad 1 \le j \le m,$$

where e_j additive noise.

Discretize by a quadrature rule:

$$\int_I \mathcal{A}(s_j, t) f(t) dt \approx \sum_{k=1}^{n} w_k \mathcal{A}(s_j, t_k) f(t_k) = \sum_{k=1}^{n} a_{jk} x_k, \quad x_k = f(t_k),$$

Discrete model in n–dimensional grid:

$$b = Ax + e, \quad A \in \mathbb{R}^{m \times n},$$

where $b_j = g(s_j)$.

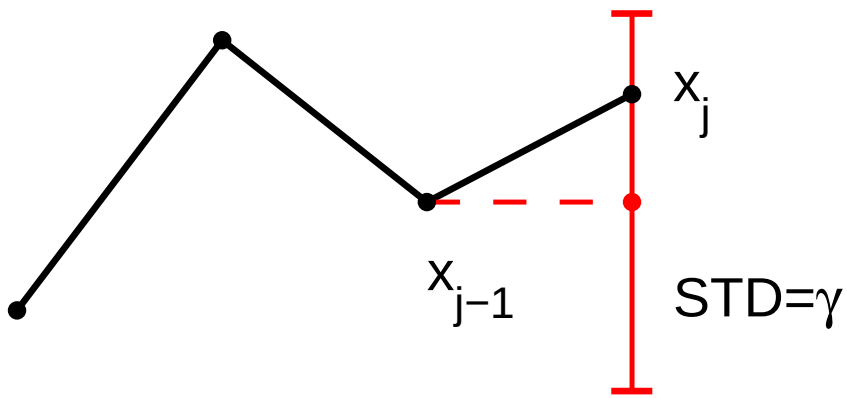
Slow variations: Markov model

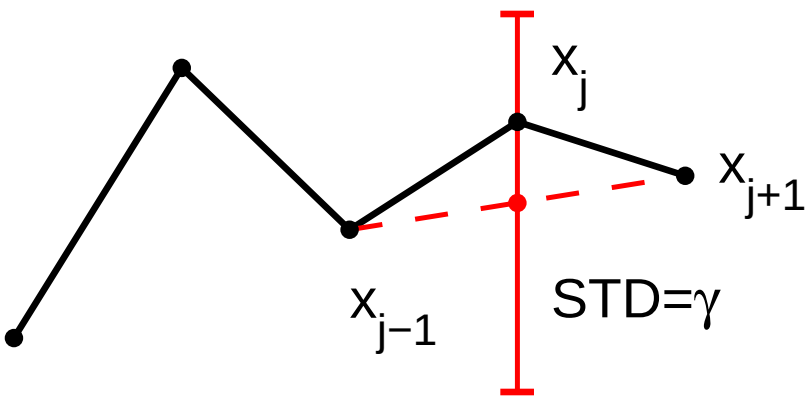
First order model:

$$X_j = X_{j-1} + \gamma W_j, \quad W_j \sim \mathcal{N}(0,1),\ W_j = \text{innovation}.$$

Second order model:

$$X_j = \frac{1}{2}(X_{j-1} + X_{j+1}) + \gamma W_j, \quad W_j \sim \mathcal{N}(0,1).$$





Matrix form

First order: $\quad L_1 X = \gamma W, \quad W \sim \mathcal{N}(0, I),$

$$L_1 = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \quad (x_0 = 0)$$

Second order: $\quad L_2 X = \gamma W, \quad W \sim \mathcal{N}(0, I),$

$$L_2 = \frac{1}{2} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ & & & -1 & 2 \end{bmatrix}, \quad (x_0 = x_{n+1} = 0)$$

Smoothness prior

The Markov model

$$LX = \gamma W, \quad W \sim \mathcal{N}(0, I), \quad (L = L_1 \text{ or } L_2)$$

implies the prior density model

$$X \sim \pi_{\text{prior}}(x) \propto \exp\left(-\frac{1}{2\gamma^2}\|Lx\|^2\right).$$

Observe: The *prior parameter* γ has an immediate interpretation, reflecting our *prior belief* of how large variations we expect.

Inhomogenous Markov model

Homogenous (or stationary) model: equal variance of every innovation step.

Inhomogenous (non-stationary) model: every innovation step may have different variance.

First order model:

$$X_j = X_{j-1} + \gamma_j W_j, \quad W_j \sim \mathcal{N}(0,1).$$

Second order model:

$$X_j = \frac{1}{2}(X_{j-1} + X_{j+1}) + \gamma_j W_j, \quad W_j \sim \mathcal{N}(0,1).$$

Inhomogenous smoothness prior

Matrix form: Define the *prior variance vector*,

$$\theta = \big[\theta_1, \theta_2, \ldots, \theta_n\big]^{\mathrm{T}}, \quad \gamma_j^2 = \theta_j.$$

$$LX = D^{1/2}W, \quad W \sim \mathcal{N}(0, I),$$

where

$$D = D_\theta = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_n).$$

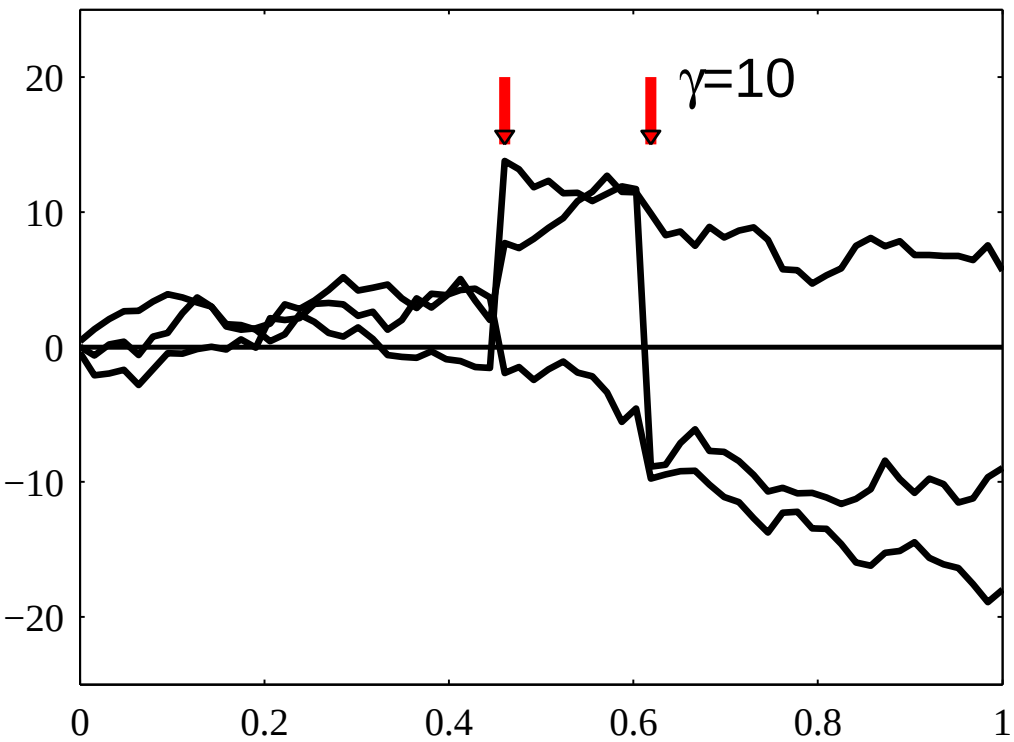
The prior density of X is then

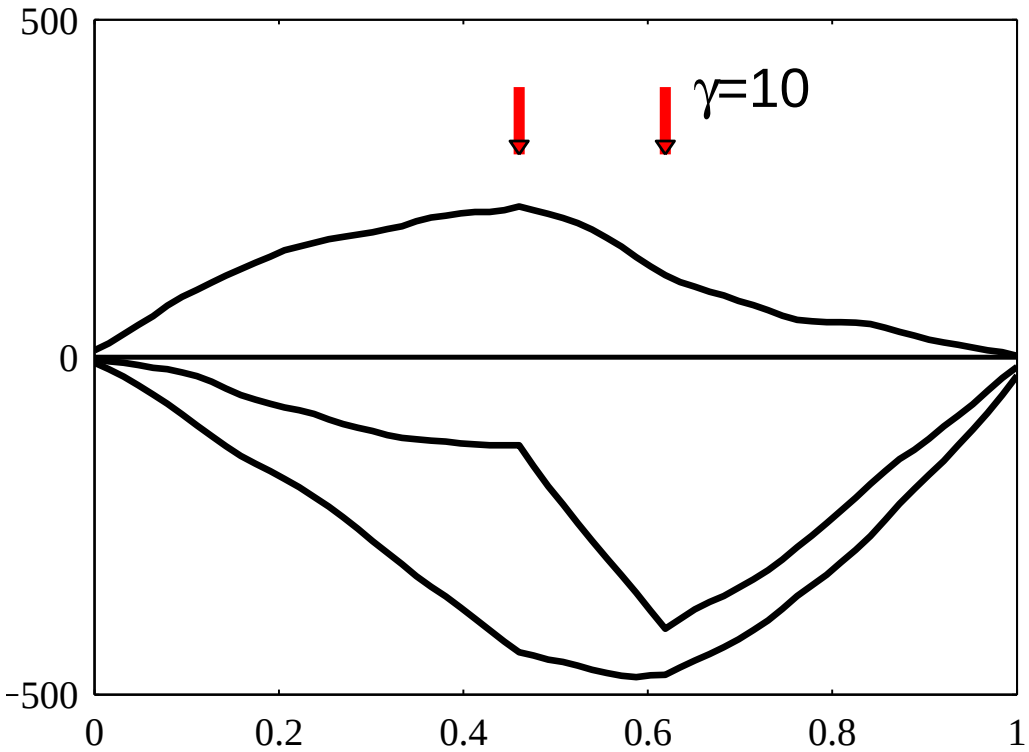
$$\pi_{\mathrm{prior}}(x) \propto \exp\left(-\frac{1}{2}\|D^{-1/2}Lx\|^2\right).$$

With different selections of θ, we can allow jumps in the solution of the inverse problem.

Testing the prior: random draws

To test whether the prior density correspond to our prior belief, draw random samples from the prior.





Solve $LX = D^{1/2}W$, where $W \sim \mathcal{N}(0, I)$.

Note: Jumps (or kinks) possible but they are not forced.

Markov model with unknown statistics

What if we don't know

- the location of the jumps,
- the size of the jumps,
- the number of the jumps?

Bayesian solution:

- Model the *prior variance vector* θ as a random variable Θ
- Estimate the pair (X, Θ) from the data and prior information.

Hierarchical model

Write a *joint prior density* for (X, Θ),

$$\pi_{\text{prior}}(x, \theta) = \pi_{\text{hyper}}(\theta)\pi_{\text{prior}}(x \mid \theta),$$

where

$$\pi_{\text{prior}}(x \mid \theta) = \left(\frac{\det(L^{\mathrm{T}} D_\theta^{-1} L)}{(2\pi)^n}\right)^{1/2} \exp\left(-\frac{1}{2}\|D_\theta^{-1/2} L x\|^2\right).$$

Observe: The normalizing constant depends on θ and cannot be ignored.

The determinant in 1D can be found explicitly. In 2D and/or with irregular grids, numerical evaluation necessary.

Do we have to compute the determinant?

No, if we pass...

... from signal to increments

Define a new random variable Z,

$$Z = LX, \quad X = L^{-1}Z,$$

and write the prior for Z:

$$\begin{aligned}
\pi_{\text{prior}}(z \mid \theta) &= \left(\frac{\det(D^{-1})}{(2\pi)^n}\right)^{1/2} \exp\left(-\frac{1}{2}\|D^{-1/2}z\|^2\right) \\
&= \left(\frac{1}{(2\pi)^n\theta_1\theta_2\cdots\theta_n}\right)^{1/2} \exp\left(-\frac{1}{2}\|D^{-1/2}z\|^2\right) \\
&= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\|D^{-1/2}z\|^2 - \frac{1}{2}\sum_{j=1}^{n}\log\theta_j\right).
\end{aligned}$$

Likelihood

Simple additive white noise model:

$$B = AX + E, \quad E \sim \mathcal{N}(0, \sigma^2 I)$$

It is assumed for simplicity that

- the noise E is independent of X (exogenous noise)
- The noise is Gaussian white noise,

$$E \sim \mathcal{N}(0, \sigma^2 I),$$

Likelihood

$$\pi(b \mid x) \exp\left(-\frac{1}{2\sigma^2}\|Ax - b\|^2\right),$$

or, in terms of the increments,

$$\pi(b \mid z) \propto \exp\left(-\frac{1}{2\sigma^2}\|AL^{-1}z - b\|^2\right).$$

Hyperprior

We expect *a priori* few jumps: the variance needs to be small except for a few outliers. Two natural candidate distributions that produce rare outliers:

Gamma distribution: $\theta_j \sim \mathrm{Gamma}(\alpha, \theta_0)$,

$$\pi_{\mathrm{hyper}}(\theta) \propto \prod_{j=1}^{n} \theta_j^{\alpha-1} \exp\left(-\frac{\theta_j}{\theta_0}\right) = \exp\left(-\frac{1}{\theta_0}\sum_{j=1}^{n}\theta_j + (\alpha-1)\sum_{j=1}^{n}\log\theta_j\right).$$

Inverse Gamma distribution: $\theta_j \sim \mathrm{InvGamma}(\alpha, \theta_0)$,

$$\pi_{\mathrm{hyper}}(\theta) \propto \prod_{j=1}^{n} \theta_j^{-\alpha-1} \exp\left(-\frac{\theta_0}{\theta_j}\right) = \exp\left(-\theta_0\sum_{j=1}^{n}\frac{1}{\theta_j} - (\alpha+1)\sum_{j=1}^{n}\log\theta_j\right).$$

Posterior density

Gamma distribution:

$$\pi(z,\theta \mid b) \propto \pi(b \mid z,\theta)\pi_{\text{prior}}(z \mid \theta)\pi_{\text{hyper}}(\theta)$$

$$\propto \exp\Big(-\frac{1}{2\sigma^2}\|AL^{-1}z-b\|^2 - \frac{1}{2}\|D^{-1/2}z\|^2 - \frac{1}{\theta_0}\sum_{j=1}^n \theta_j + \Big(\alpha - \frac{3}{2}\Big)\sum_{j=1}^n \log\theta_j \Big).$$

Inverse Gamma distribution:

$$\pi(z,\theta \mid b) \propto \pi(b \mid z,\theta)\pi_{\text{prior}}(z \mid \theta)\pi_{\text{hyper}}(\theta)$$

$$\propto \exp\Big(-\frac{1}{2\sigma^2}\|AL^{-1}z-b\|^2 - \frac{1}{2}\|D^{-1/2}z\|^2 - \theta_0\sum_{j=1}^n \frac{1}{\theta_j} - \Big(\alpha + \frac{3}{2}\Big)\sum_{j=1}^n \log\theta_j \Big).$$

Algorithms

Several possibilities:

- Compute the MAP estimator $(z_{\mathrm{MAP}}, \theta_{\mathrm{MAP}})$,
- Compute the marginal density of θ maximize the *evidence*,

$$\theta_* = \mathrm{argmax} \int_{\mathbb{R}^n} \pi(z, \theta \mid b) dz,$$

  and estimate then z from $\pi(z, \theta_* \mid b)$,
- Compute the Posterior Mean estimate, e.g., by MCMC methods.
- Estimate the reliability of the point estimates estimate by MCMC.

The first one is fast, the last one in necessary for reliability assessment.

Fast track: MAP estimate

MAP estimate $(x_{\mathrm{MAP}}, \theta_{\mathrm{MAP}})$ is the minimizer of the negative of the log-posterior. With Gamma distribution,

$$F(z,\theta) = \frac{1}{2\sigma^2}\|AL^{-1}z - b\|^2 + \frac{1}{2}\|D_\theta^{-1/2}z\|^2 + \frac{1}{\theta_0}\sum_{j=1}^n \theta_j - \left(\alpha - \frac{3}{2}\right)\sum_{j=1}^n \log\theta_j.$$

Quick iterative solver (cyclic coordinate iteration):

1. Initialize $\theta = \theta^0$, $k = 1$.
2. Update z,
$$z^k = \mathrm{argmin} F(z, \theta^{k-1}).$$
3. Update θ,
$$\theta^k = \mathrm{argmin} F(z^k, \theta).$$
4. Increase k by one and repeat from 2. until convergence.

This algorithm is fast because ...

1. Updating w:

$$z^k = \text{argmin}\left(\frac{1}{2\sigma^2}\|AL^{-1}z - b\|^2 + \frac{1}{2}\|D^{-1/2}z\|^2\right), \quad D = D_{\theta^{k-1}},$$

that is, z^k is the least squares solution of the system

$$\begin{bmatrix} (1/\sigma)AL^{-1} \\ D^{-1/2} \end{bmatrix} z = \begin{bmatrix} (1/\sigma)b \\ 0 \end{bmatrix}.$$

2. Updating θ: θ_j^k satisfies

$$\frac{\partial}{\partial\theta_j}F(z^k,\theta) = -\frac{1}{2}\left(\frac{z_j^k}{\theta_j}\right)^2 + \frac{1}{\theta_0} - \left(\alpha - \frac{3}{2}\right)\frac{1}{\theta_j} = 0,$$

which has an explicit solution,

$$\theta_j^k = \theta_0\left(\eta + \sqrt{\frac{(z_j^k)^2}{2\theta_0} + \eta^2}\right), \quad \eta = \frac{1}{2}\left(\alpha - \frac{3}{2}\right).$$

Gamma distribution and Total Variation

Total Variation (TV) regularization:[5]

Minimize

$$\int_I |\mathcal{A}f(t) - g(t)|^2 dt + \kappa \int_I |Df(t)| dt.$$

Euler-Lagrange equation

$$\mathcal{A}^*(\mathcal{A}f - g) + \kappa \mathrm{D}^* \left(\frac{1}{|\mathrm{D}f|} \mathrm{D}f \right) = 0$$

with homogenous Neumann boundary conditions.

[5] Rudin L I, Osher S and Fatemi E: Nonlinear total variation based noise removal algorithms. Physica D **60** 259–68.

Fixed point iteration[6]:

Given the approximate solution f^{k-1}, new iterate f^k by solving

$$\mathcal{A}^*(\mathcal{A}f - g) + \kappa \mathrm{D}^* \left(\frac{1}{|\mathrm{D}f^{k-1}|} \mathrm{D}f \right) = 0. \tag{1}$$

Finite difference approximation: $\mathrm{D} \to (T/n)L$, and $\mathrm{D}^* \to (T/n)L^{\mathrm{T}}$,

$$A^{\mathrm{T}}(Ax - b) + \left(\frac{n\kappa}{T} \sqrt{\frac{\theta_0}{2}} \right) L^{\mathrm{T}} \big(D_{\theta^{k-1}}^{-1} Lx \big) = 0, \quad \theta^{k-1} = |x_j^{k-1} - x_{j-1}^{k-1}|,$$

which are the normal equations of the LSQR problem, with $\sigma^2 = n\kappa/T\sqrt{\theta_0/2}$.

[6] Vogel C R and Oman M E: Fast, robust total variation-based reconstruction of noisy, blurred images. IEEE Trans. Image Process. **7** (1998) 813–824.

Inverse Gamma and Perona-Malik

MAP estimation algorithm with Inverse Gamma corresponds to the fixed point iteration of the equation

$$\mathcal{A}^*(\mathcal{A}f - g) + \kappa \mathrm{D}^* \left(\frac{1}{1 + \beta |\mathrm{D}f|^2} \mathrm{D}f \right) = 0,$$

where

$$\kappa = \sigma^2 T^2 (\alpha + 3/2)/(n^2 \theta_0), \qquad \beta = T^2/(2\theta_0 n^2).$$

This is in fact the *Perona-Malik functional*[7].

The Perona-Malik functional is also closely related to the *Mumford-Shah functional*[8].

[7] Perona P and Malik J: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12** (1990) 629–639.

[8] Mumford D and Shah J: Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. **42** (1989) 577–684.

Statistical Reliability Analysis

Markov Chain Monte Carlo (MCMC): Generate a large sample

$$\{(z^1, \theta^1), (z^2, \theta^2), \ldots, (z^N, \theta^N)\}$$

that is distributed according to the posterior density.

Block form Gibbs sampler:

1. Initialize by selecting (z^1, θ^1) and set $k = 1$.
2. Draw z^{k+1} from the distribution $z \mapsto \pi(z \mid \theta^k) \propto \pi_{\text{post}}(z, \theta^k)$.
3. Draw θ^{k+1} from the distribution $\theta \mapsto \pi(\theta \mid z^{k+1}) \propto \pi_{\text{post}}(z^{k+1}, \theta)$.
4. Unless $k = N$, increase k by one and repeat from 2.

In Practice

Updating of z:

$$\pi(z \mid \theta^k) \propto \exp\left(-\frac{1}{2\sigma^2}\|b - AL^{-1}z\| - \frac{1}{2}\|D^{-1/2}z\|^2\right), \quad D = \mathrm{diag}(\theta^k),$$

is Gaussian: solve

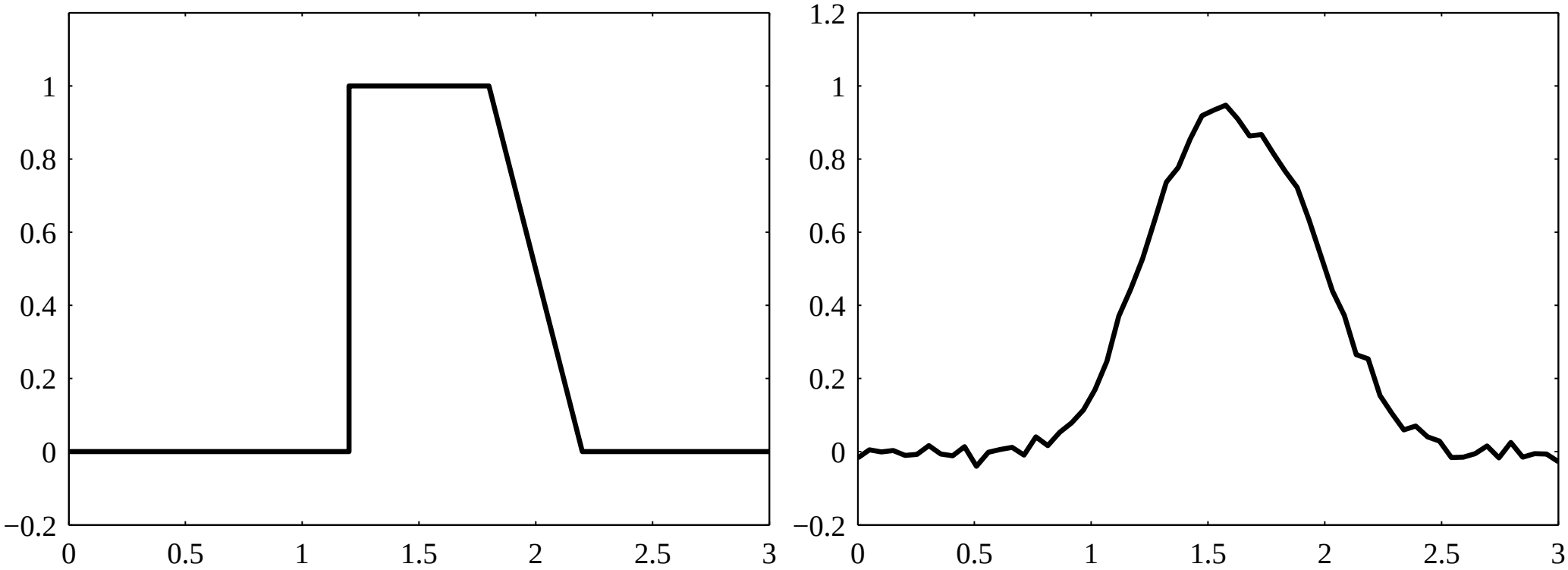
$$\begin{bmatrix} (1/\sigma)AL^{-1} \\ D^{-1/2} \end{bmatrix} z = \begin{bmatrix} (1/\sigma)b \\ 0 \end{bmatrix} + w, \quad w \sim \mathcal{N}(0, I),$$

where $w \in \mathbb{R}^{m+n}$ is a realization of the white noise process.

The updating of θ can be done componentwise: Draw θ_j from the one-dimensional density

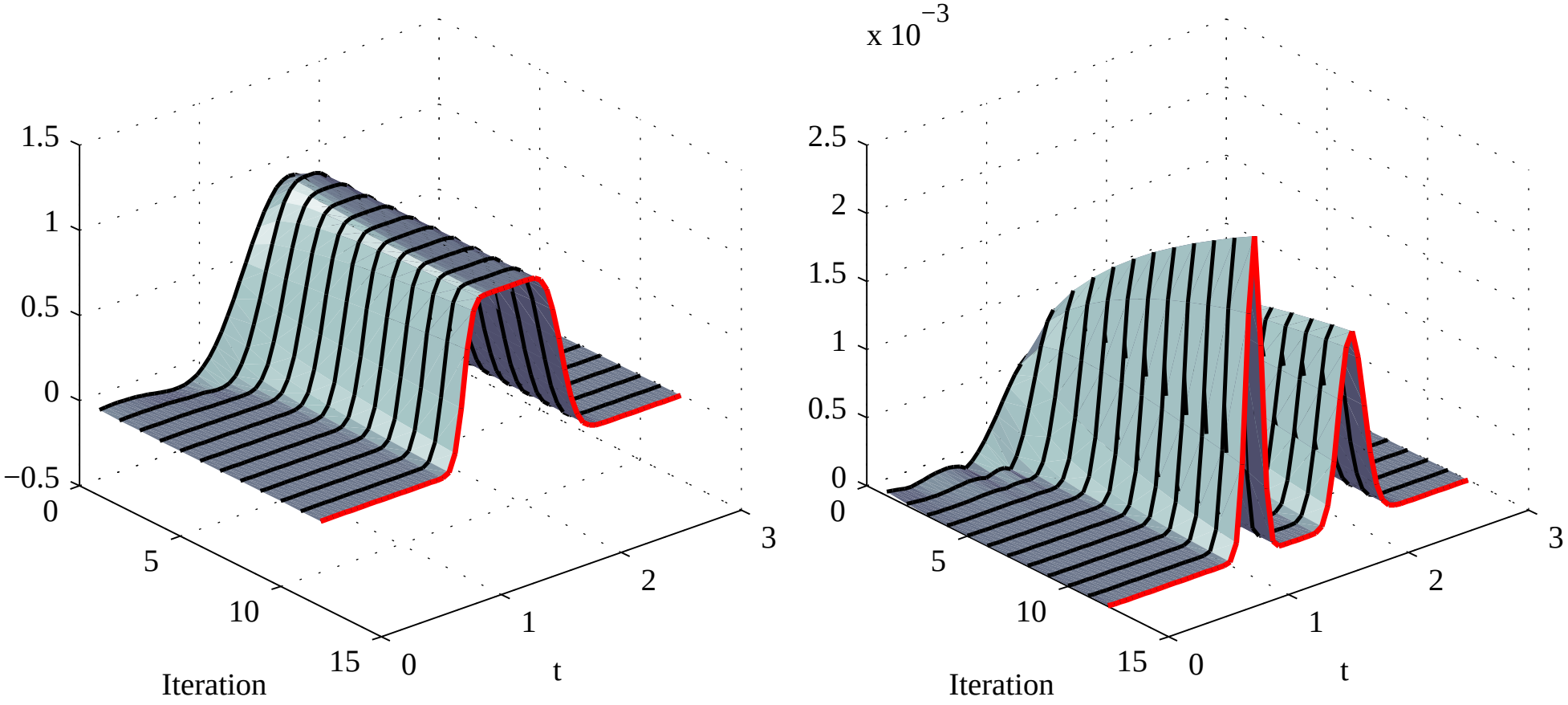
$$\theta_j \mapsto \exp\left(-\frac{(z_j^{k+1})^2}{2\theta_j} - \frac{\theta_j}{\theta_0} + \left(\alpha - \frac{3}{2}\right)\log\theta_j\right).$$

Computed example: Signal and data

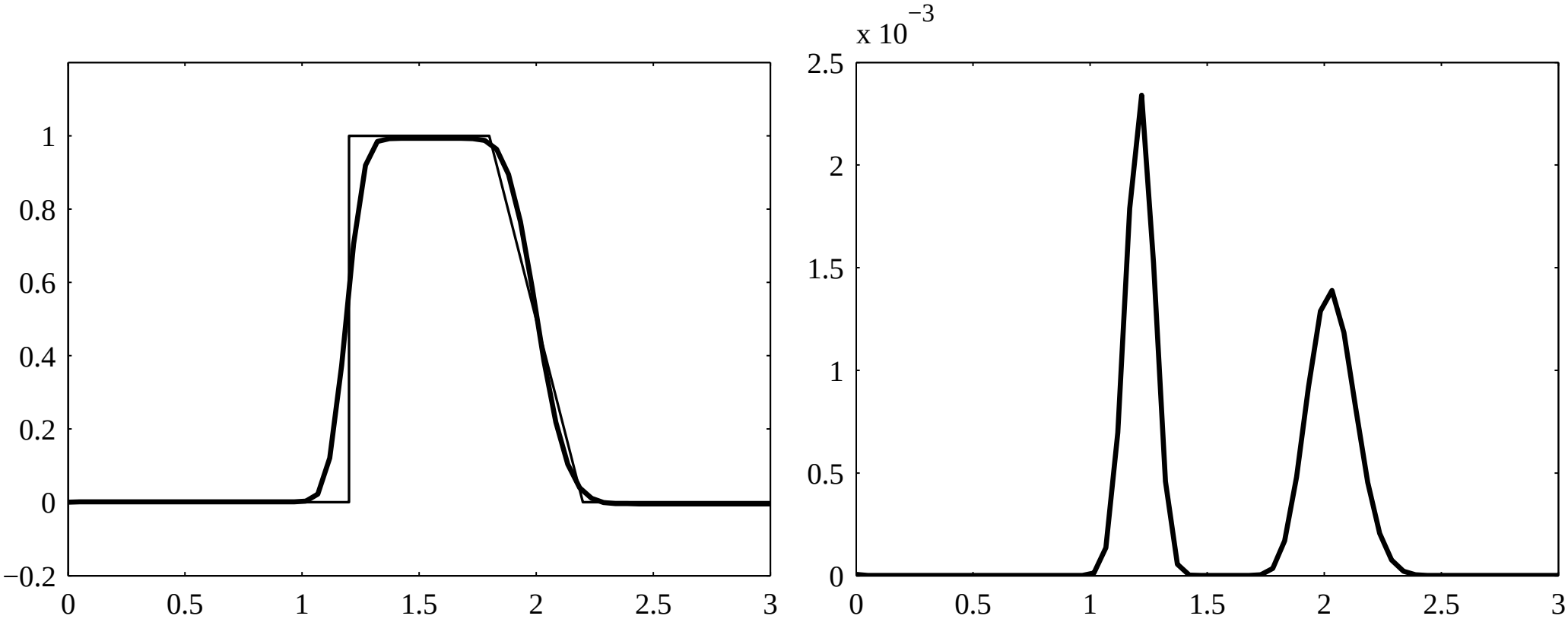


Gaussian blur, noise level 2%

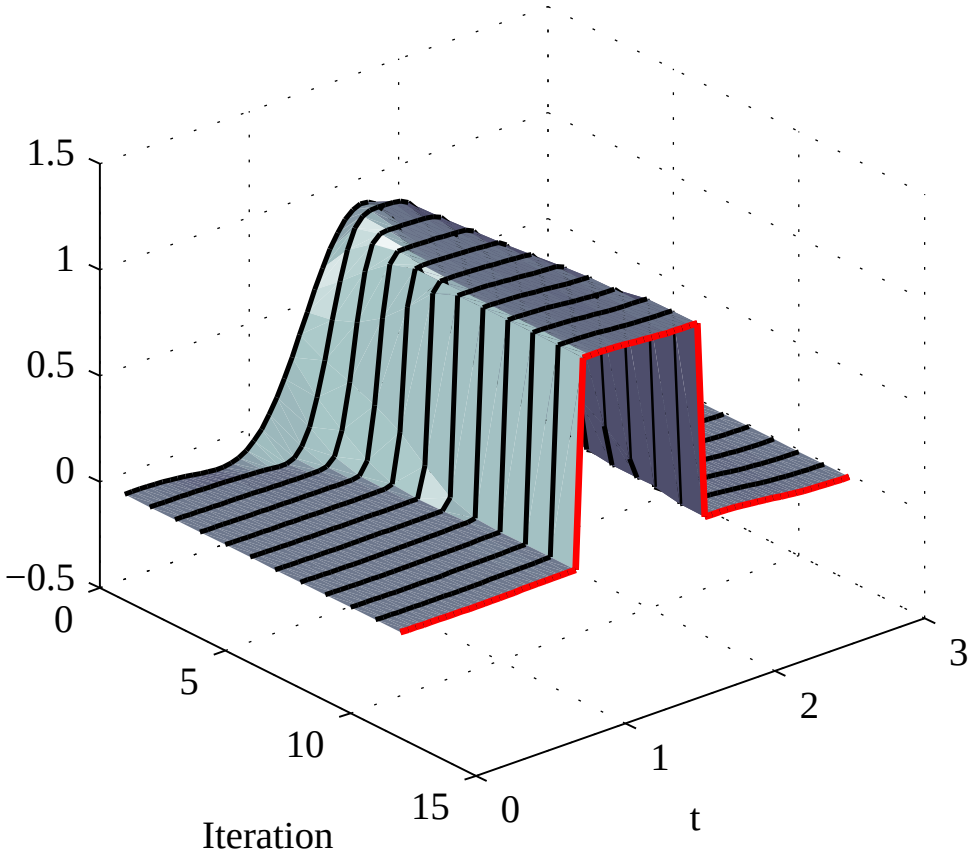
MAP estimate, Gamma hyperprior

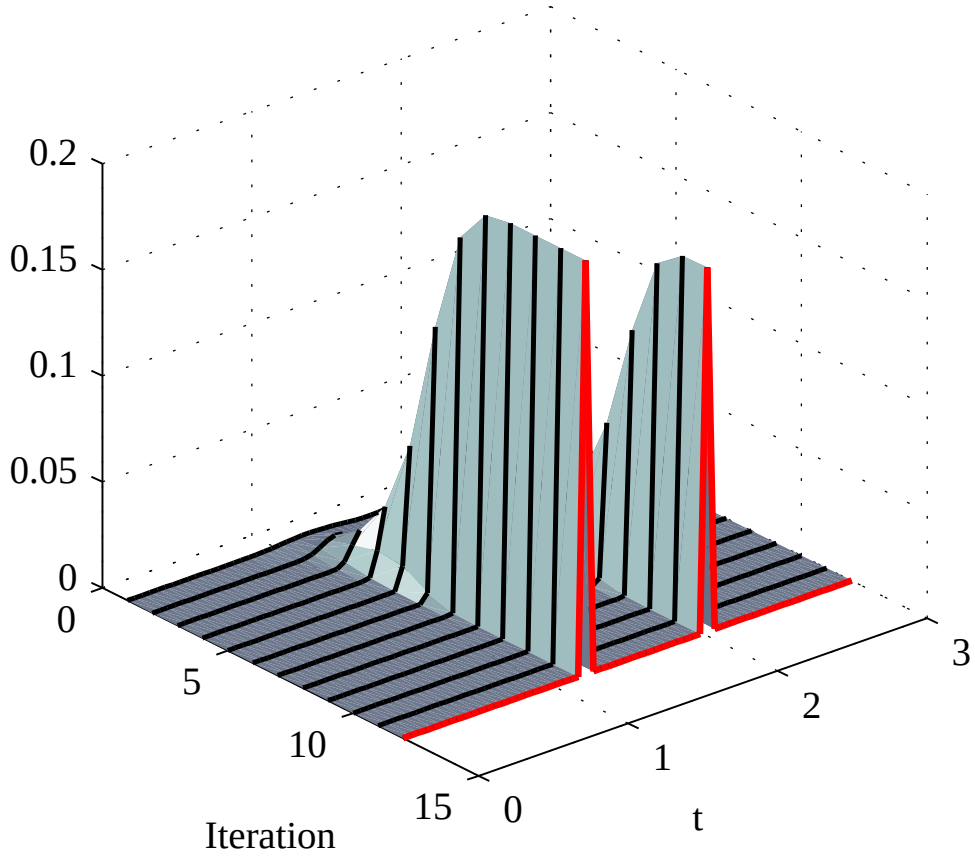


MAP estimate, Gamma hyperprior

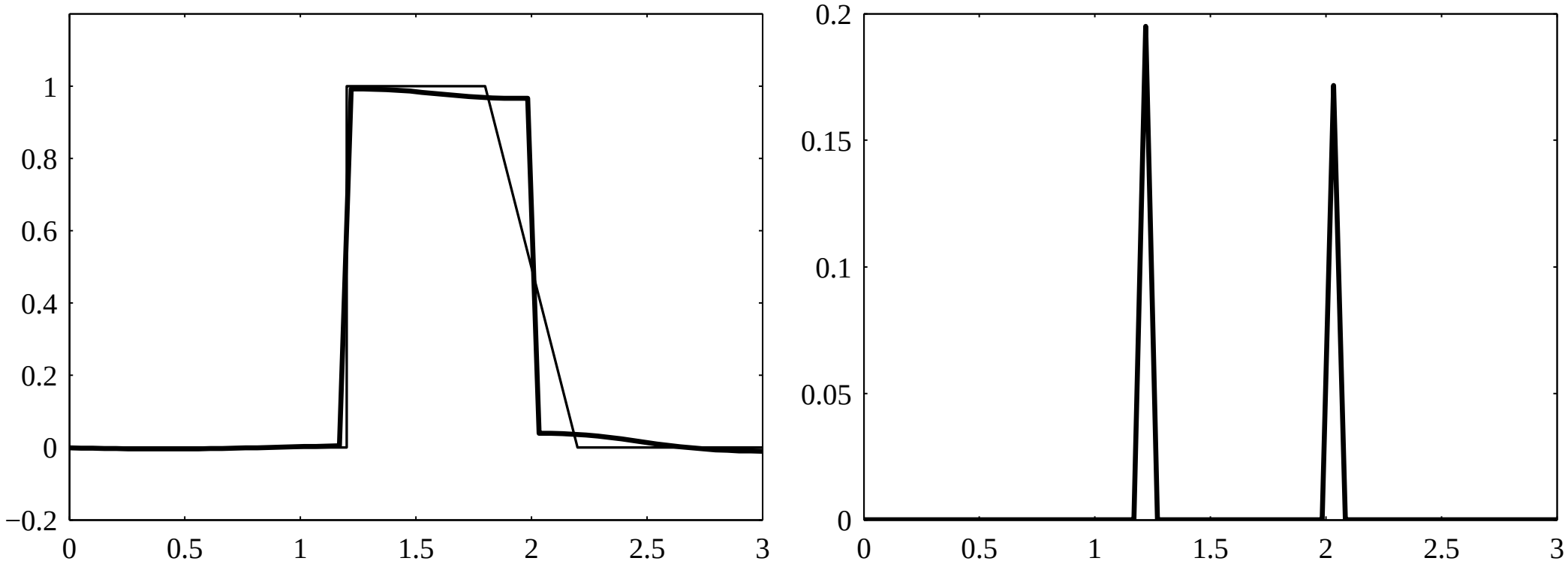


MAP estimate, Inverse Gamma hyperprior

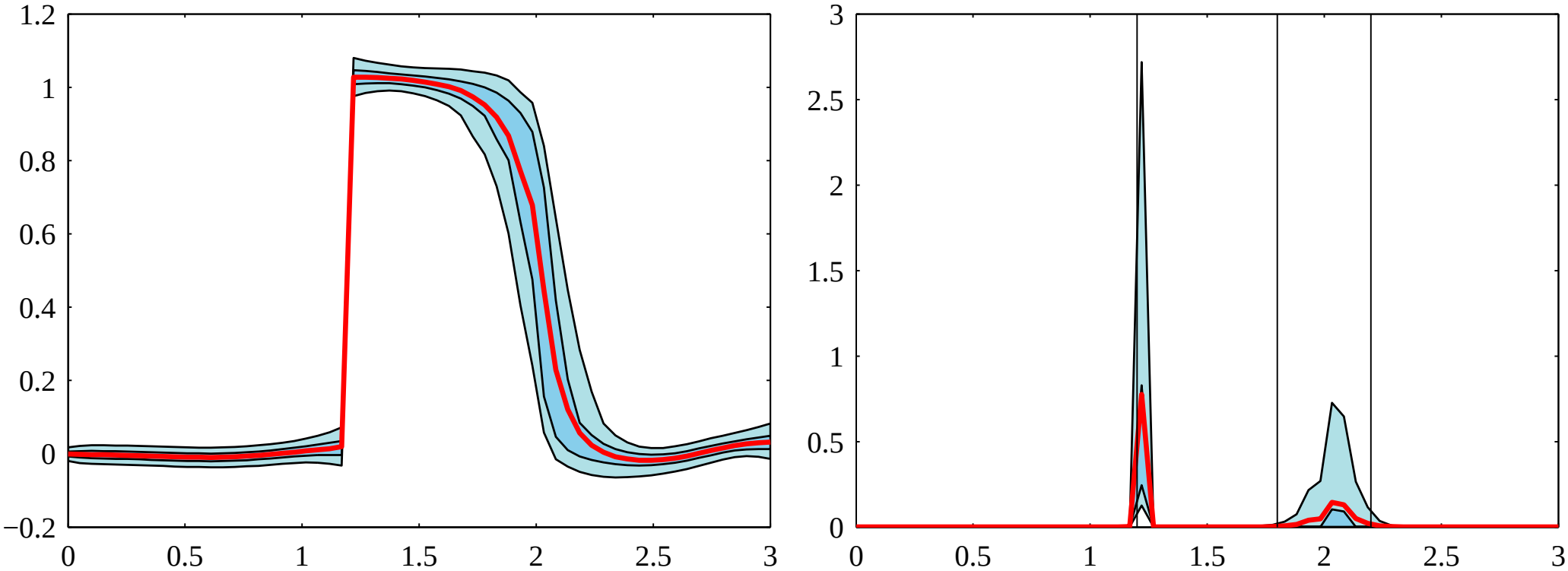




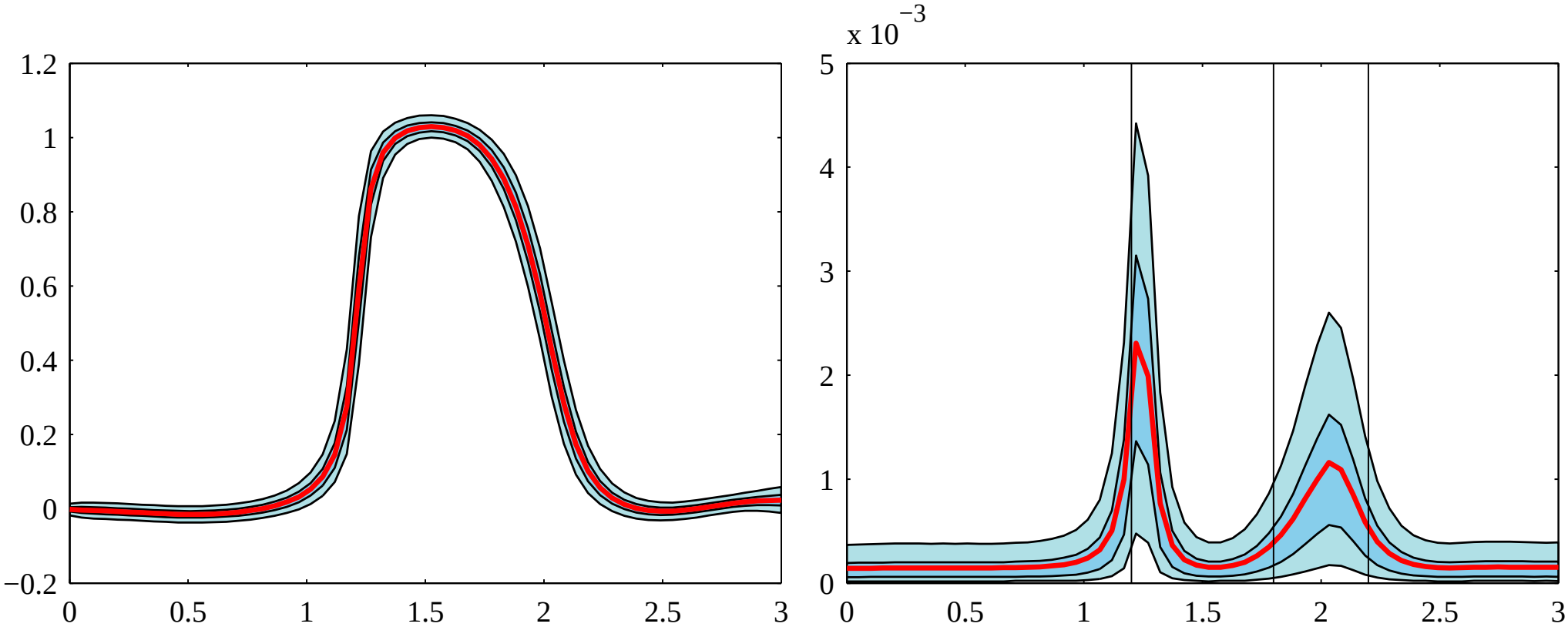
MAP estimate, Inverse Gamma hyperprior



MCMC predictive output envelopes, Inverse Gamma hyperprior



MCMC predictive output envelopes, Gamma hyperprior



Extension in two dimensions

Special attention to

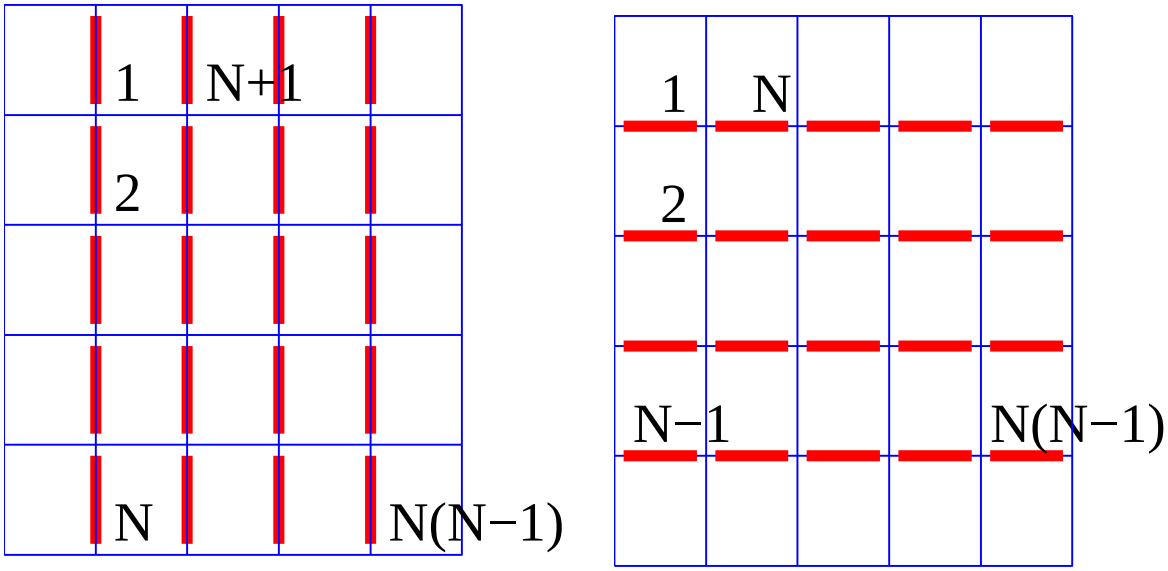
- How to define the increment vector in the Markov model,
- Image quality, e.g., rotational invariance,
- Memory requirements,
- Computational cost.

Markov model: Count of variables

Number of pixels $N_{\mathrm{p}} = N^2$.

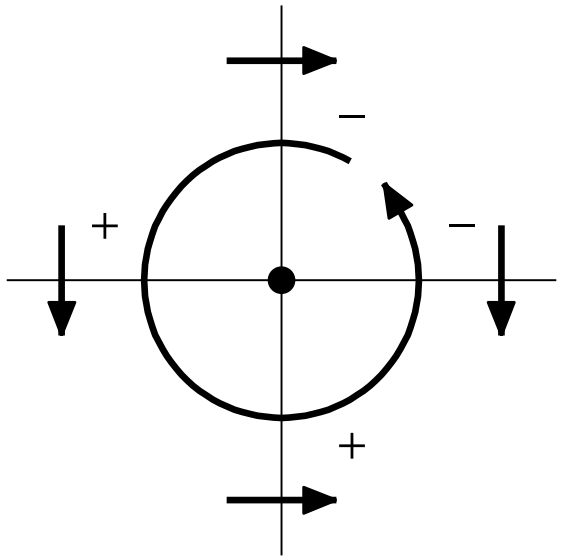
Number of increments (edges) $N_{\mathrm{e}} = 2N(N-1) > N_{\mathrm{p}}$.

Hence, no one-to-one correspondence.



Compatibility conditions

The increments are not independent:



Number of interior nodes $N_{\mathrm{n}} = (N-1)^2$.

Count independent edges: $2N(N-1) - (N-1)^2 = N^2 - 1 = N_{\mathrm{p}} - 1$.

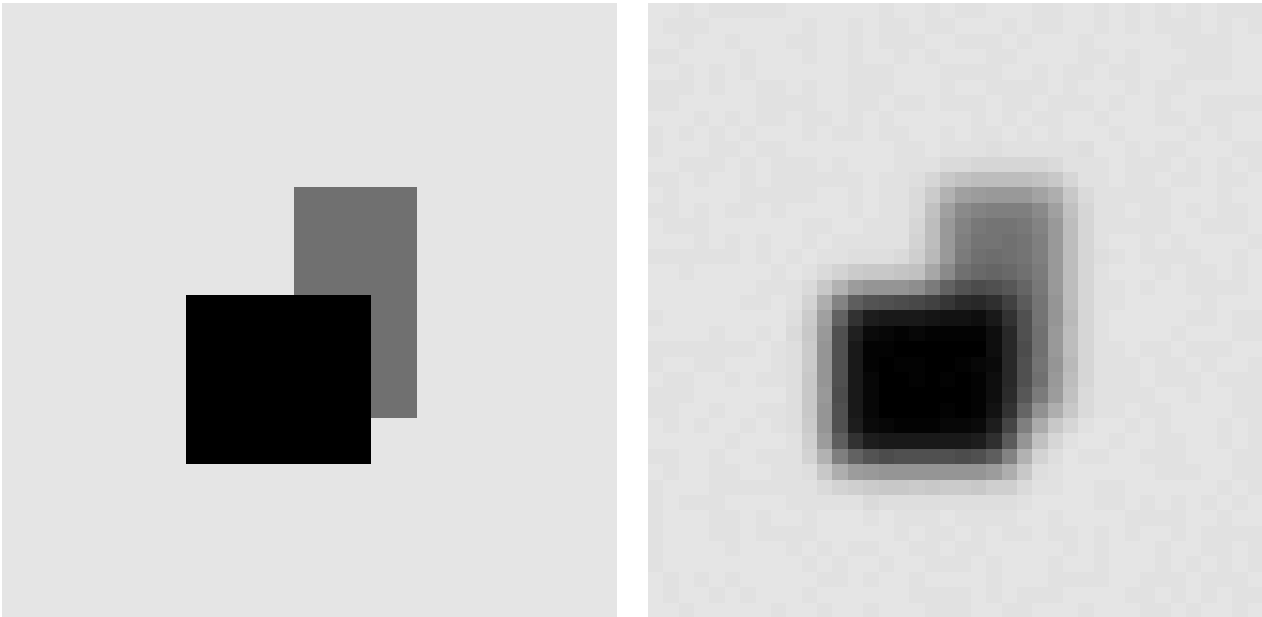
We need one extra condition, e.g., the average of the image.

How to build the algorithm?

1. Define $z \in \mathbb{R}^{N_e}$, the vector of the increments over the edges,
2. Define $\theta \in \mathbb{R}^{N_e}$, the variances of the increments,
3. Define one extra degree of freedom $z_{N_e+1} \in \mathbb{R}$, e.g., the average of the image,
4. Using the *compatibility condition* $Mz = 0$, express the pixel image in terms of z,
5. Write a hypermodel,
6. Solve the MAP estimate by alternating iteration.

Skipping the details...

Deblurring of a simple image

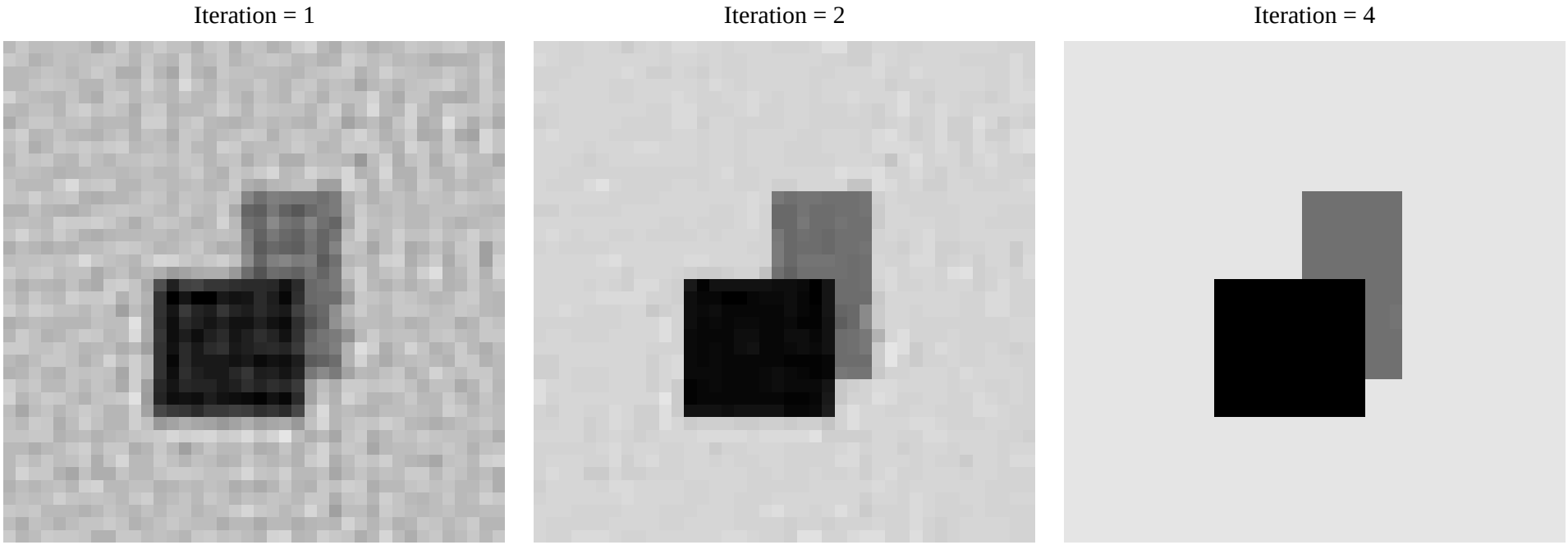


Original image 40×40

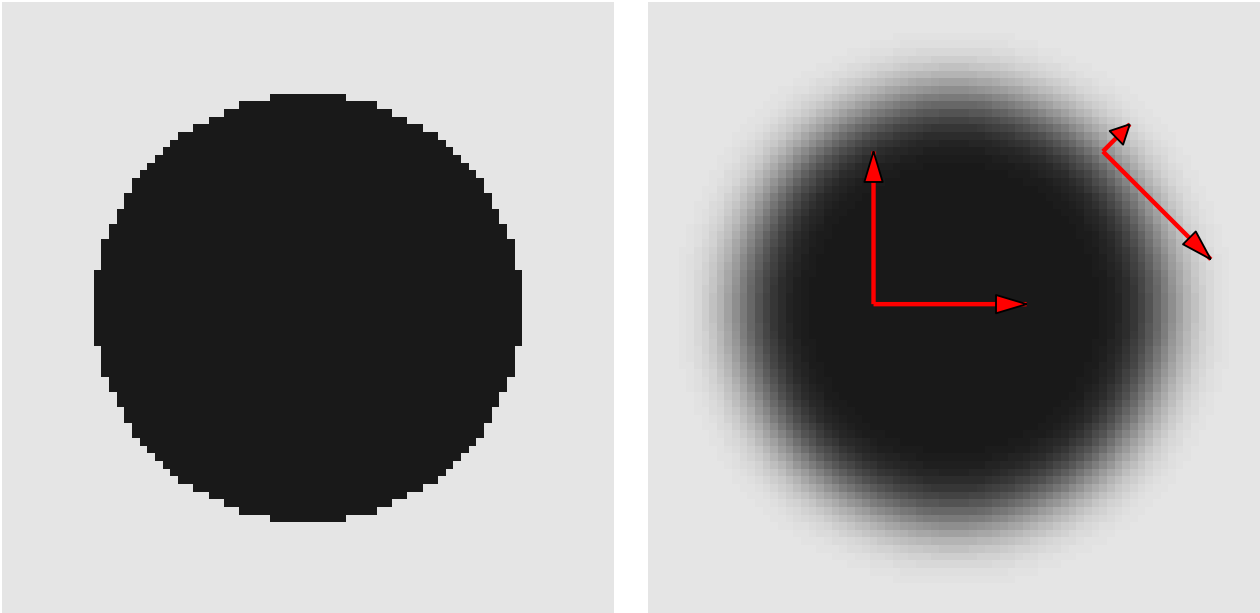
Blurring with a Gaussian kernel, FWHV = 3 pixels

Additive Gaussian white noise, STD = 0.5% of max. of the noiseless signal.

A few iterations



Bayesian learning, bootstrap prior[9]



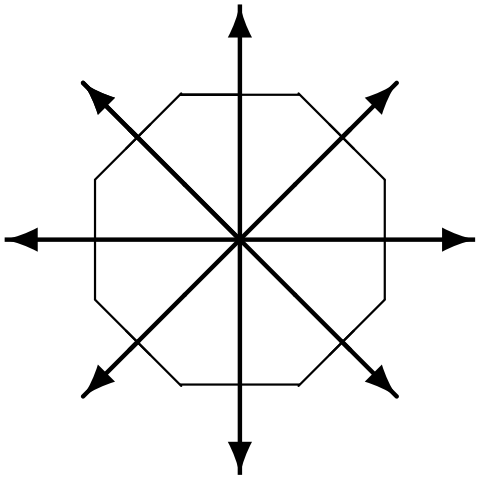
Update your current information based on the latest reconstruction. May look dubious, but is justified by the hierarchical modelling:

$$\pi_{\mathrm{pr}}^k(z) = \pi_{\mathrm{pr}}(z \mid \theta^k), \quad \theta^k = g(z^k).$$

[9] D. Calvetti and E. Somersalo: Microlocal sequential regularization in imaging. Inverse Problems and Imaging **1** (2007) 1–11.

Rotational invariance

The algorithm works well with blocky structures that have edges parallel to the coordinate axes. For oblique edges, staircasing may occur.



A better mask for finite difference approximations needed.

Diffusion operator

Let $u = u(r)$ be a differentiable vector field in $\mathbb{R}^2$.

By the Divergence Theorem,

$$\begin{aligned}
\nabla \cdot u(x) &= \lim_{\varepsilon \to 0} \frac{1}{\pi \varepsilon^2} \int_{\{|r-r'|<\varepsilon\}} \nabla \cdot u(r') dr' \\
&= \lim_{\varepsilon \to 0} \frac{1}{\pi \varepsilon^2} \int_{\{|r-r'|=\varepsilon\}} n(r') \cdot u(r') dS(r') \\
&= \lim_{\varepsilon \to 0} \frac{1}{\pi \varepsilon} \int_{\mathbb{S}^1} \theta \cdot u(r + \varepsilon\theta) d\theta \\
&= \lim_{\varepsilon \to 0} M_\varepsilon u(r).
\end{aligned}$$

Approximate

$$\nabla \cdot \big(\lambda \nabla f\big) \approx M_\varepsilon \big(\lambda \nabla f\big) = \frac{1}{\pi \varepsilon} \int_{\mathbb{S}^1} \lambda(r + \varepsilon\theta) \theta \cdot \nabla f(r + \varepsilon\theta) d\theta.$$

Microlocal model

Replace $\lambda(r+\varepsilon\theta)$ by a non-linear functional of f,

$$\lambda = \lambda\big[f\big](r,\theta) = \frac{1}{1+(\tau|\theta\cdot\nabla f(r)|)^\kappa}, \quad \tau>0,\ \kappa>0.$$

Algorithm:

1. Initialize $f_0 = 1$ and $k=0$.
2. Define

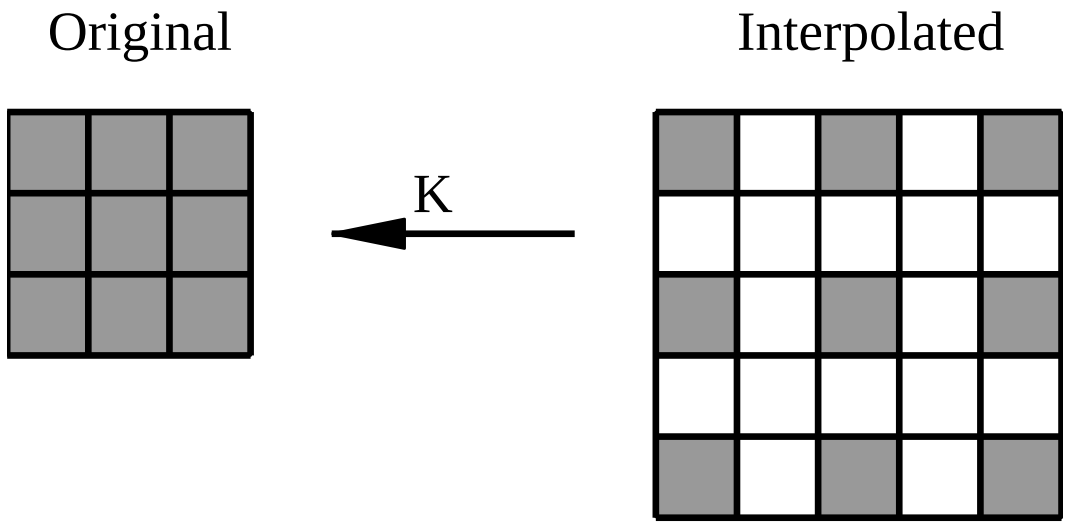
$$f_{k+1} = \arg\min\left\{\int_{Q'} |\mathcal{A}f(s)-g(s)|^2 ds + \gamma\int_Q |M_\varepsilon\big(\lambda\big[f_k\big]\nabla f(r)\big)|^2 dr\right\},$$

   where

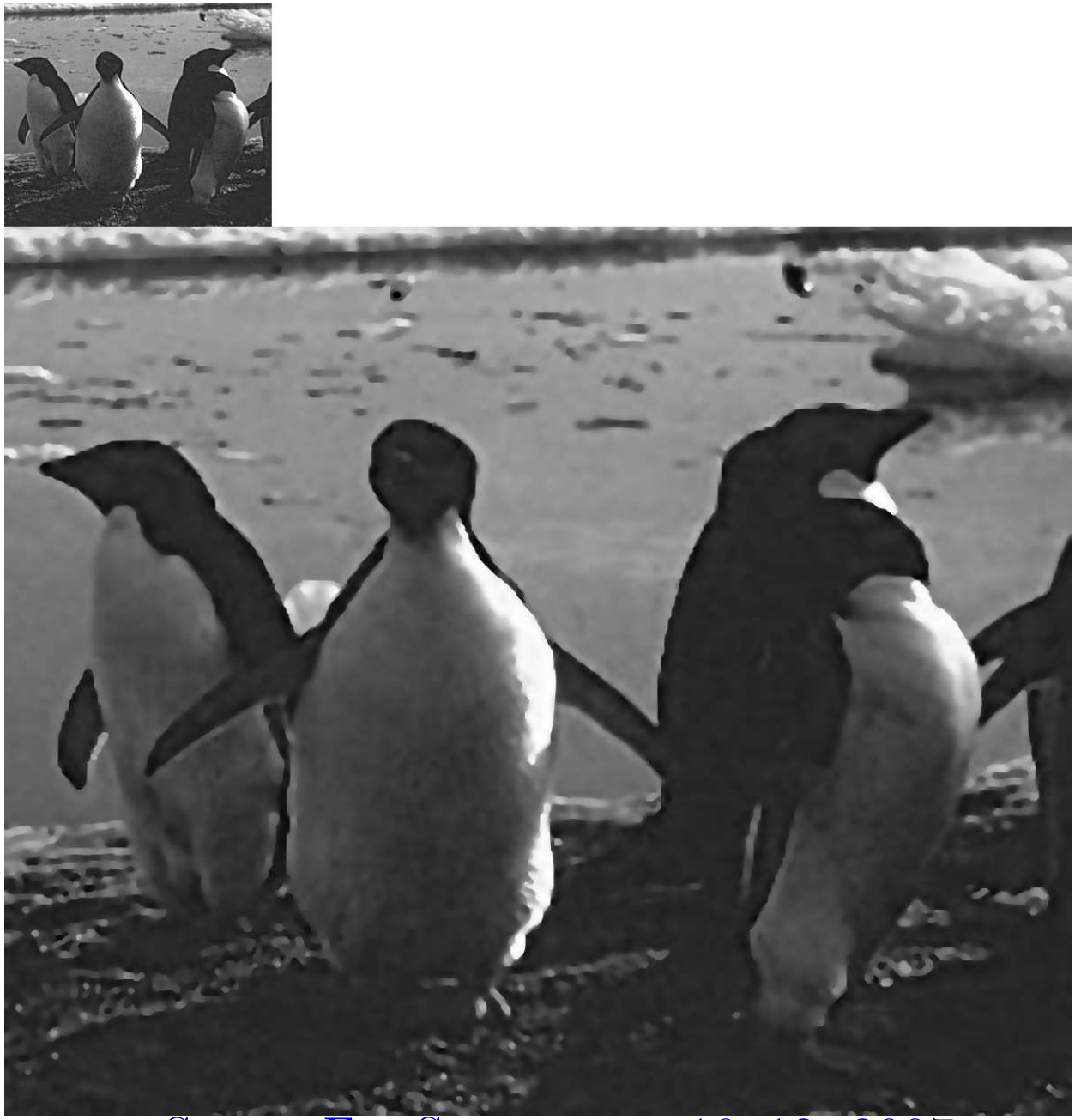
$$\lambda\big[f_k\big](r,\theta) = \frac{1}{1+(\tau|\theta\cdot\nabla f_k(r)|)^\kappa}, \quad \tau>0,\ \kappa>0.$$

3. Increase k by one and repeat from 2 until a stopping criterion is satisfied.

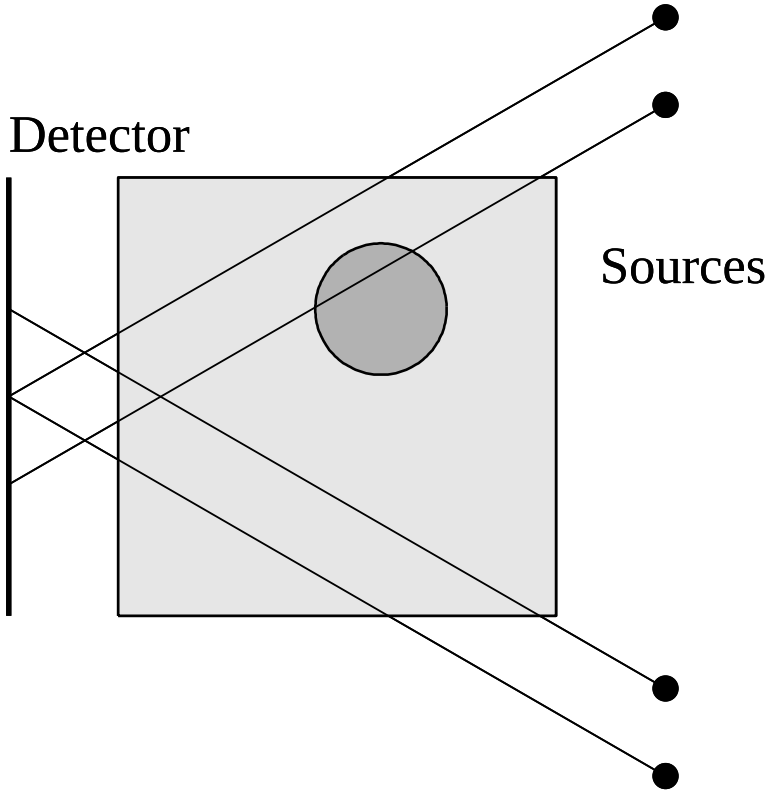
Example: image zooming

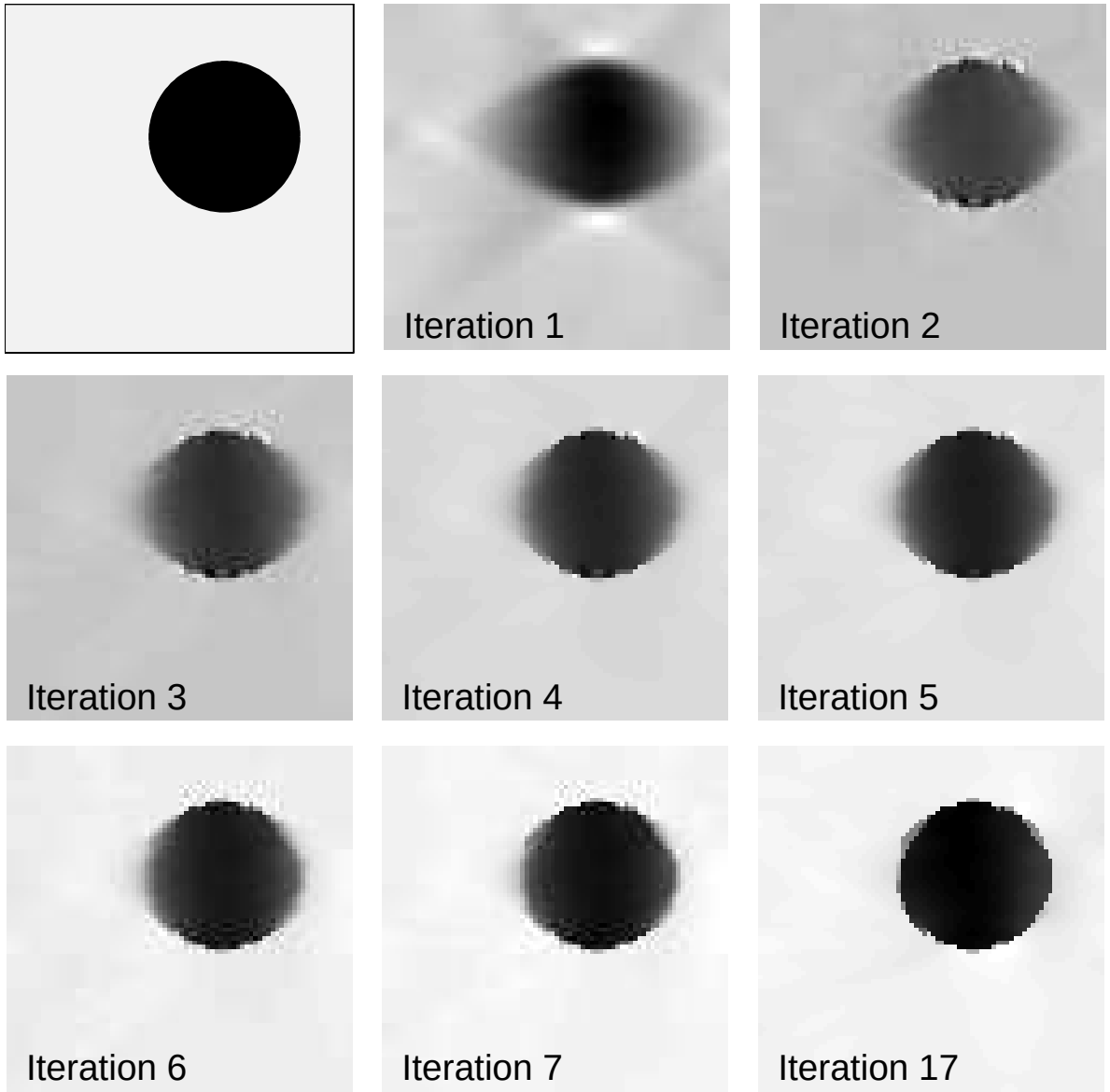


Forward map is the sparse sampling matrix.

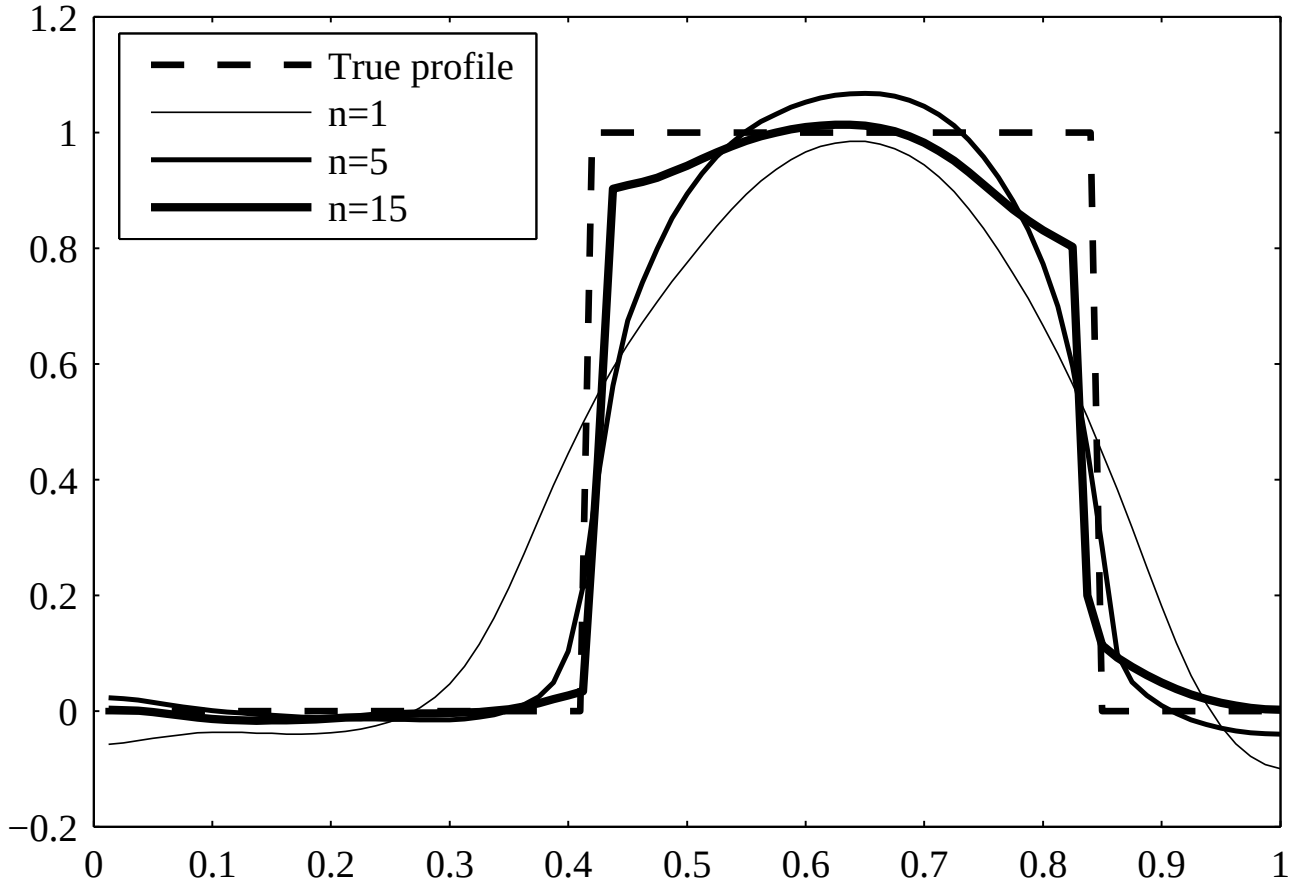


Limited angle tomography




Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5
Iteration 6
Iteration 7
Iteration 17

Sectional profiles



Speeding up: iterative solvers

At each iteration step, we need to minimize

$$x^{k+1} = \text{argmin}(\|Ax - b\|^2 + \|L^k x\|^2), \quad L^k = L[x^k].$$

- Memory and computational costs require iterative solvers
- In practice, CGLS is the only method of choice
- CGLS is equivalent to solve the normal equations
- Since A^{T} is smoothing, to get sharp edges, an enormous amount of iterations are required

Way out: priorconditioners

Whitening: Write

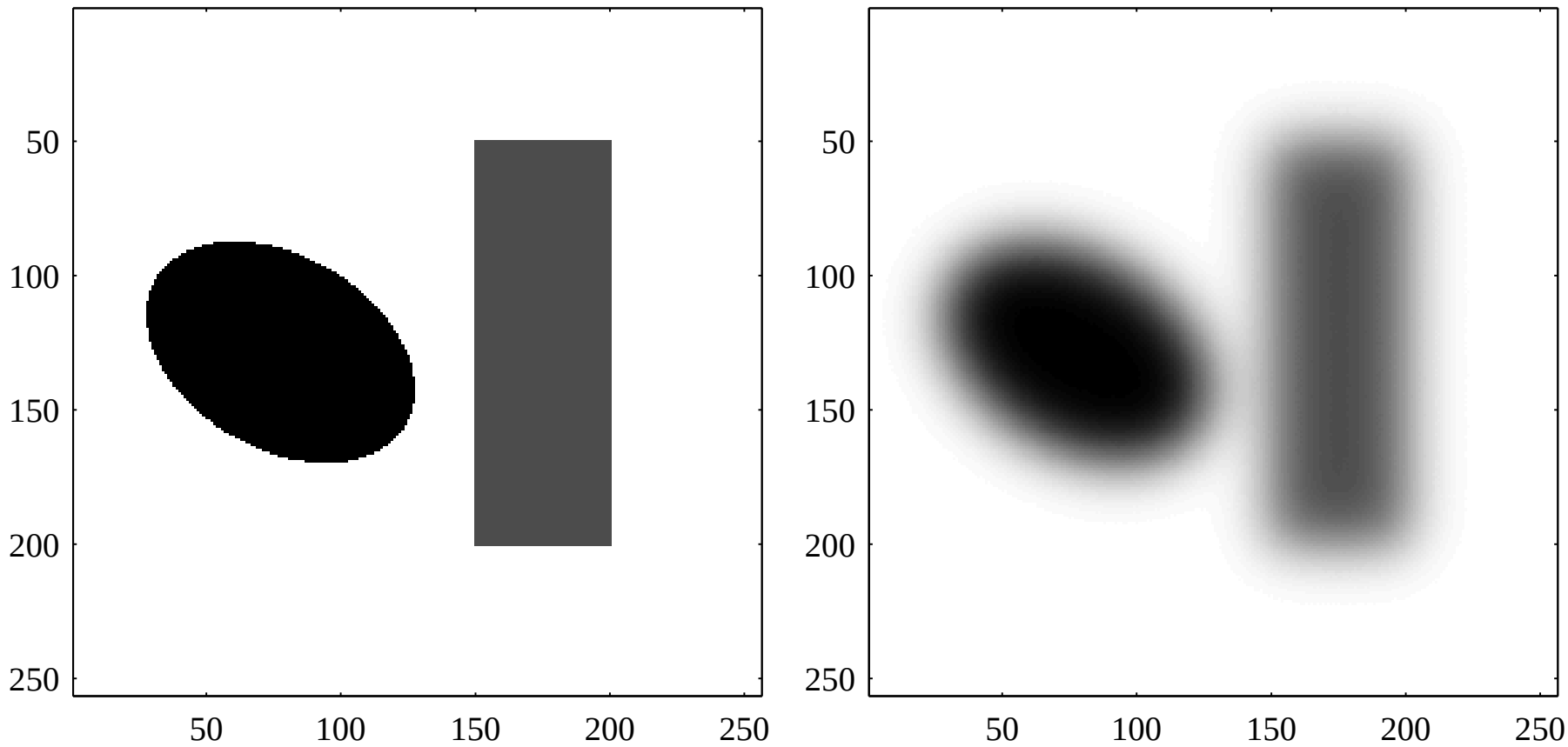
$$\|A(L^k)^{-1}w - b\|^2 + \|w\|^2, \quad L^k x = w,$$

Instead of solving the original LSQR problem use GMRES to the *right preconditioned* system

$$A(L^k)^{-1}w = b, \quad L^k x = w,$$

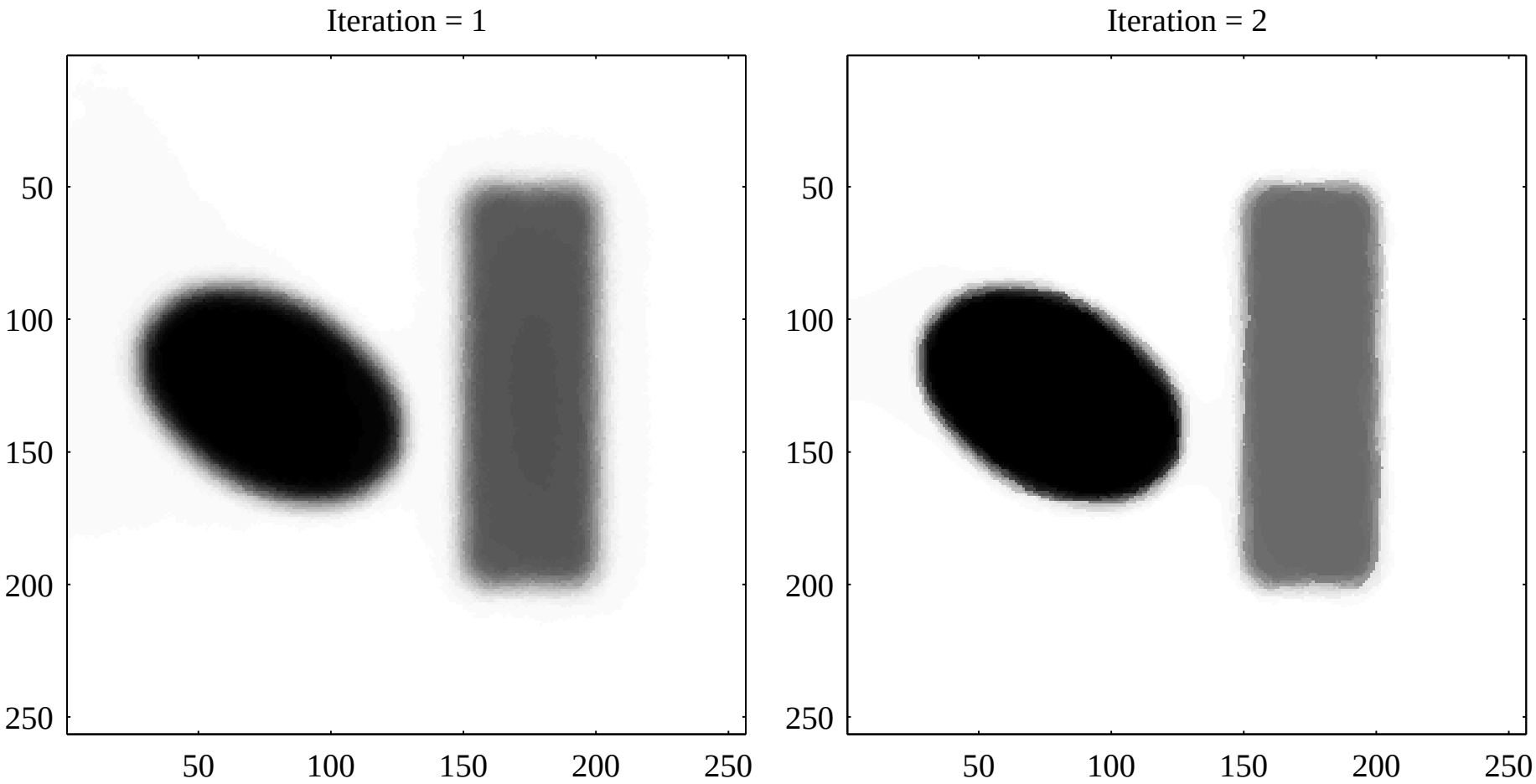
regularized by early truncation of the iterations.

Deblurring example



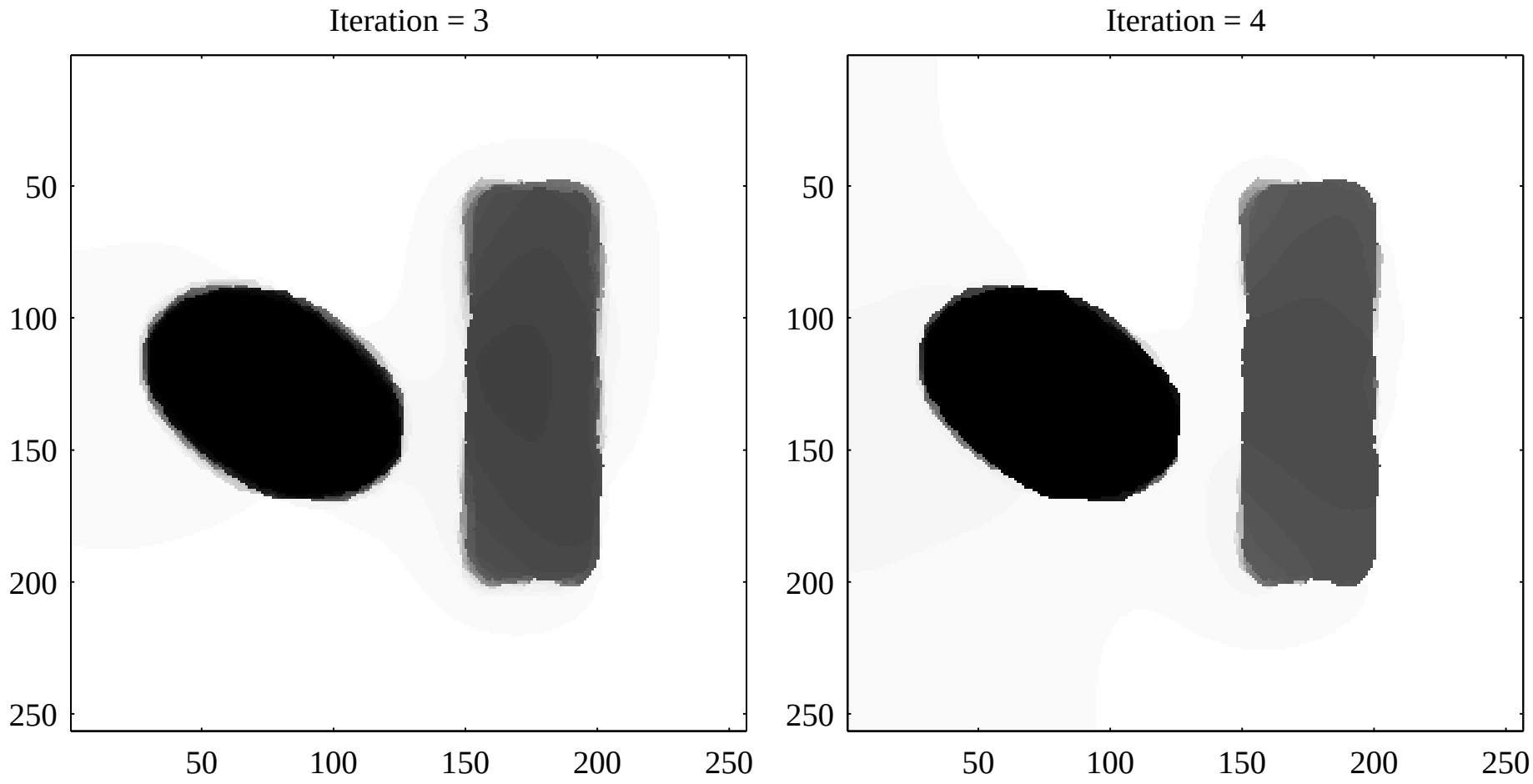
True and blurred image

Deblurring example

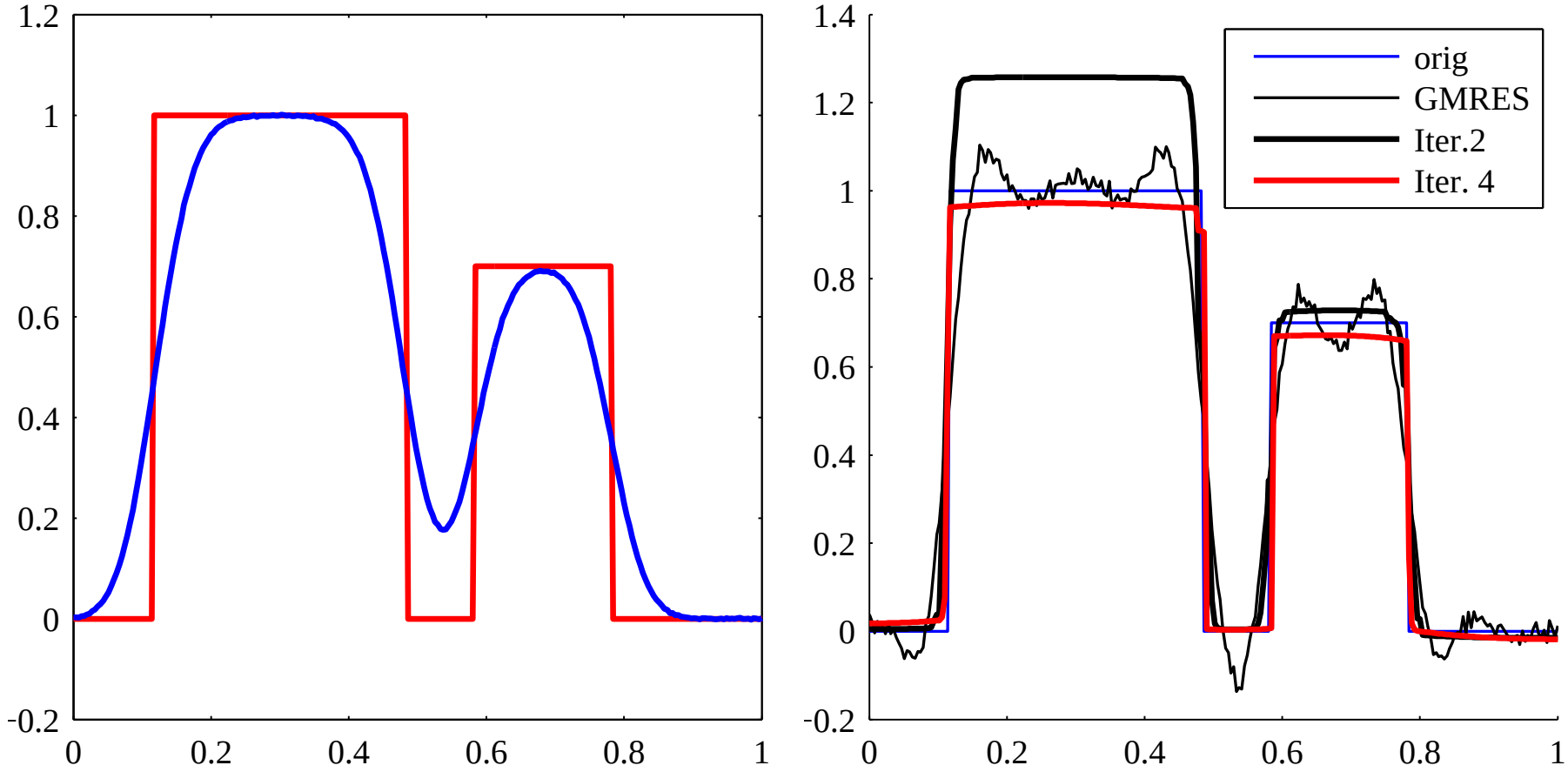
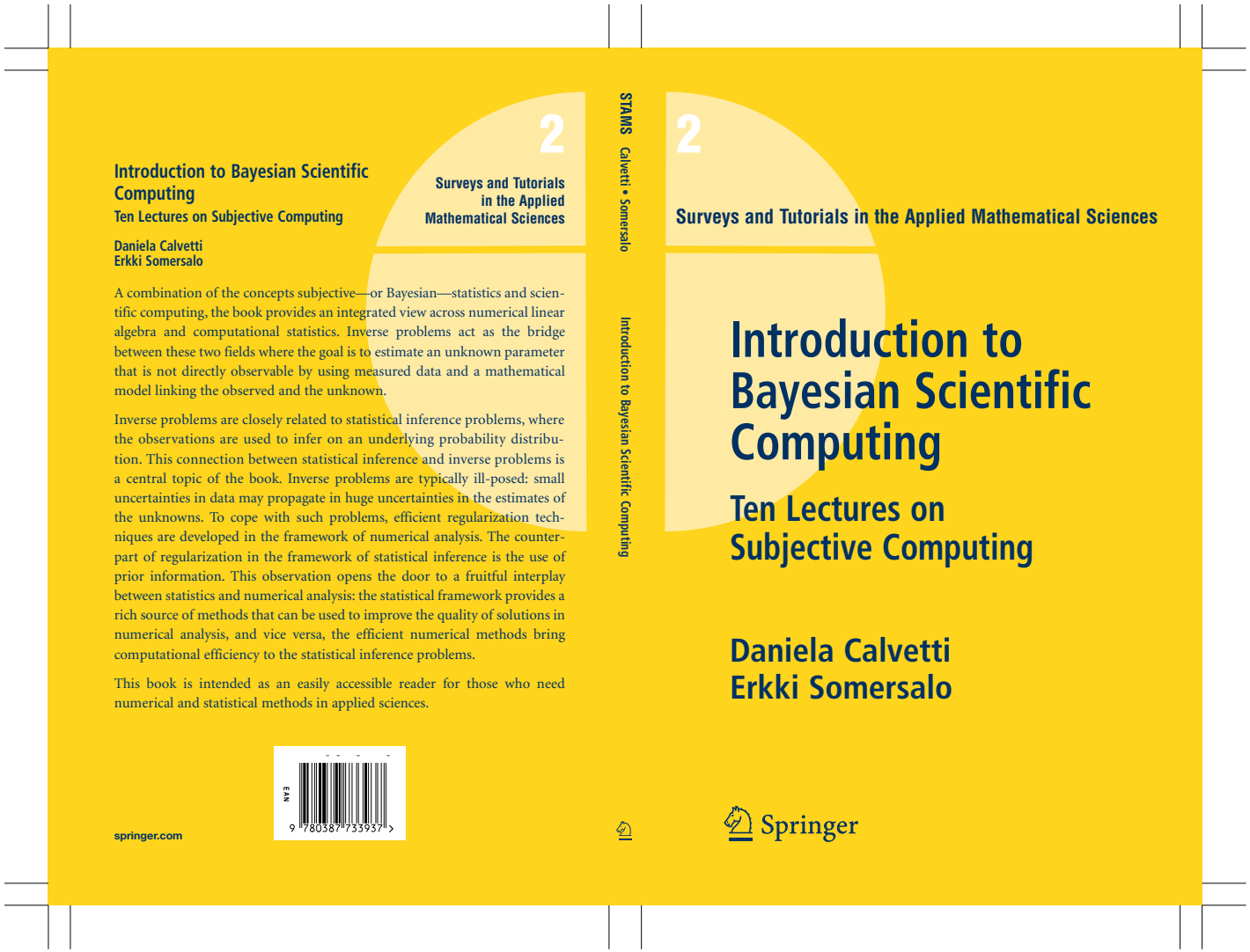


Priorconditioned GMRES, 18 seconds/iteration on a 2GB laptop in Matlab.

Deblurring example



Deblurring example



More on these topics, see ...