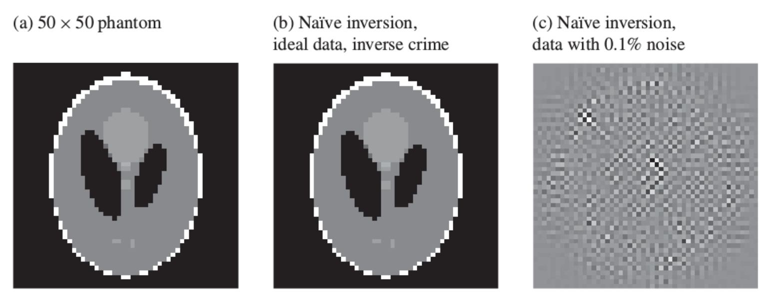
Basics of X-ray CT: Noise

Viktoria Markova

Sarntal, some September 2018



The problem

The Problem: Find f

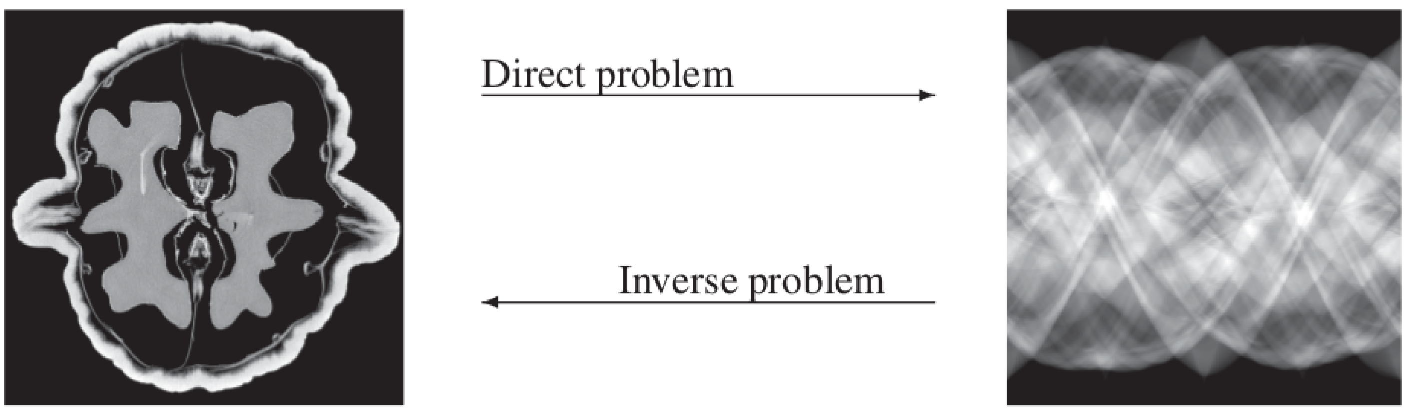
find me

$$\mathbf{m} = A\, f + \varepsilon$$

noise

measurement

linear operator

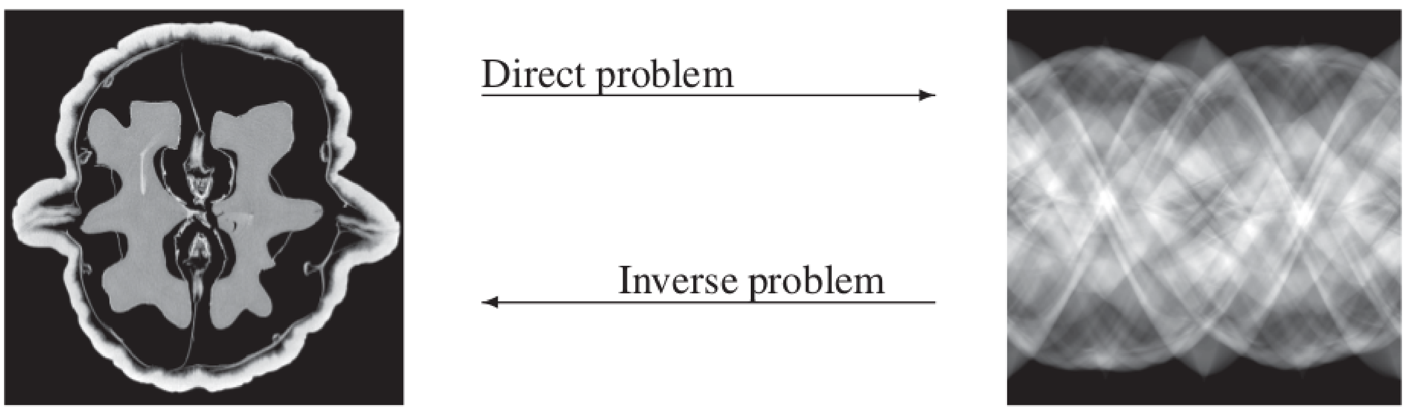


The problem

The Problem: Find f

$$\mathbf{m} = R\, f + \varepsilon$$

- f: find me
- ε: noise
- $\mathbf{m}$: measurement
- R: linear operator: the Radon transform



The problem

Naive reconstruction:

$$f \approx A^{-1}\mathbf{m} \tag{1}$$

The problem

Naive reconstruction: *"inverse crime"*

$$f \approx A^{-1}\mathbf{m} \tag{1}$$

(a) 50 × 50 phantom

(b) Naïve inversion, ideal data, inverse crime

(c) Naïve inversion, data with 0.1% noise

What do we do?

Impose Occam's razor by imposing a prior distributions.

What do we do?

Impose Occam's razor by imposing a prior distributions.

Aha, right. So what do we actually do?

Overview

Overview I

Tikhonov regularization

- Generalized version
- Computation
- Parameter choice
  - Morozov discrepancy principle
  - L-curve method

Overview II

Total variation regularization

- Comparison to second norm (Tikhonov reg.)
- Computational approaches
  - Quadratic programming
  - Large-scale gradient-based minimization method

Overview III

Curvelet Sparse Regularization

- Curvelet frame
- Computation with ADMM
  - Parameter discussion
- Comparison to TV

Tikhonov regularization

- First choice for linear problems + generalized form accommodates the usage of known properties
- Not edge preserving

$$v = \underset{\mathbf{z} \in \mathbb{R}^n}{\arg\min} ||A\mathbf{z} - \mathbf{m}||^2 + \alpha ||\mathbf{z}||^2 \tag{2}$$

- Small residual $A\mathbf{v} - \mathbf{m}$
- v small in L^2 norm (prevents overfitting)

Generalized Tikhonov regularization (priori information)

- f is close to f_*:
$$v = \arg\min ||A\mathbf{z} - \mathbf{m}||^2 + \alpha||\mathbf{z} - f_*||^2 \quad (3)$$
- f is known to be smooth:
$$v = \arg\min ||A\mathbf{z} - \mathbf{m}||^2 + \alpha||L\mathbf{z}||^2 \quad (4)$$
  or
$$v = \arg\min ||A\mathbf{z} - \mathbf{m}||^2 + \alpha||L(\mathbf{z} - f_*)||^2 \quad (5)$$
  L is a discretized differential operator/matrix

Tikhonov regularization: Computation

- Stacked form of the non-generalized equation:

$$\begin{bmatrix} A \\ \sqrt{\alpha} \end{bmatrix} f = \begin{bmatrix} \mathbf{m} \\ 0 \end{bmatrix} \tag{6}$$

written as

$$\widetilde{A}f = \widetilde{m} \tag{7}$$

leading to the solution by computing the least-square (no need to compute the SVD):

$$f = \widetilde{A} \backslash \widetilde{m} \tag{8}$$

Generalized form:

$$v = (A^T A + \alpha L^T L)^{-1} A^T \mathbf{m} \tag{9}$$

Compute with the *conjugate gradient method*. (No need to construct the matrices A, A^T, L, L^T)

Tikhonov regularization

- Simple implementation
- Problem: How to choose parameter?

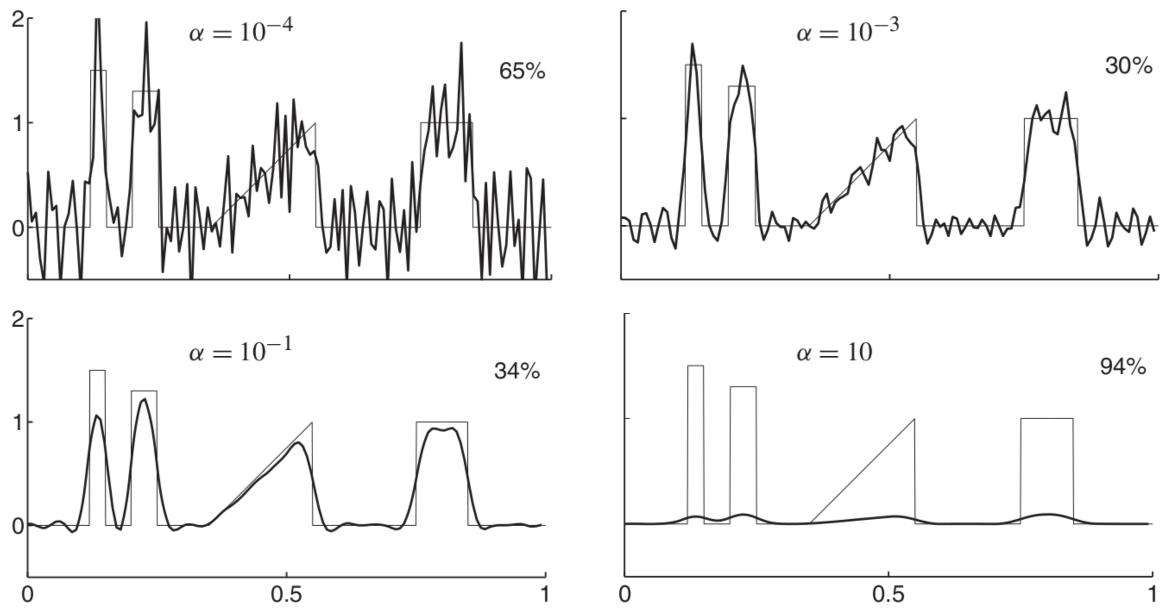


Figure 5.1. Tikhonov regularized reconstructions. The percentages shown are relative errors of reconstructions.

Tikhonov regularization: Parameter choice

- Morozov discrepancy principle
- L-curve method
- And other methods e.g. Generalized cross-validation method

Tikhonov regularization: Parameter choice

- **Morozov discrepancy principle**
- L-curve method
- And other methods e.g. Generalized cross-validation method

Parameter choice: Morozov discrepancy principle

Estimate on error exists => solution with the same level of noise is ok, so choose α, such that

$$||Av - \mathbf{m}|| = noise \tag{10}$$

If

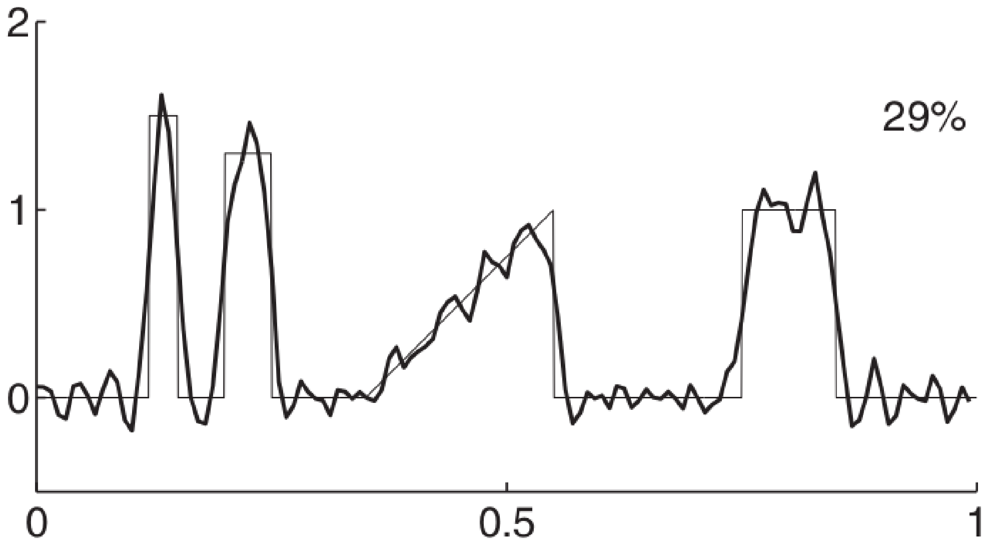
$$||P\mathbf{m}|| <= noise <= ||\mathbf{m}|| \tag{11}$$

then α is unique.

P - orthogonal projection to the subspace $Coker(A)$

Result

After a simple computation of a longer formula for the deconvolution problem with $noise = 11\%$:



Problem

Morozov discrepancy principle does not apply to the generalized version.

Tikhonov regularization: Parameter choice

- Morozov discrepancy principle
- **L-curve method**
- And other methods e.g. Generalized cross-validation method

Parameter choice: L-curve method

1. Gather candidates for α
2. Compute the result v for each.
3. Plot $log(||Av - \mathbf{m}||), log(||Lv||)$ for results
4. Observe the "L curve graph". "Optimal" solution at the corner.

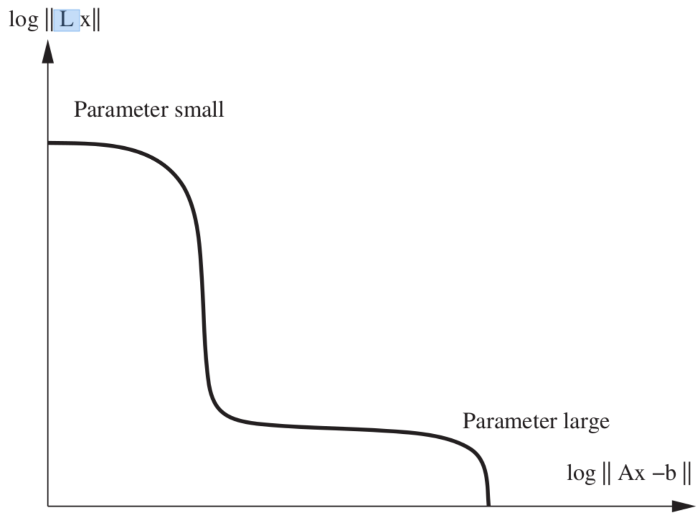


Figure 5.8. An idealized illustration of an L-curve formed by plotting a continuum of points defined by (5.19).

L-curve method: Example

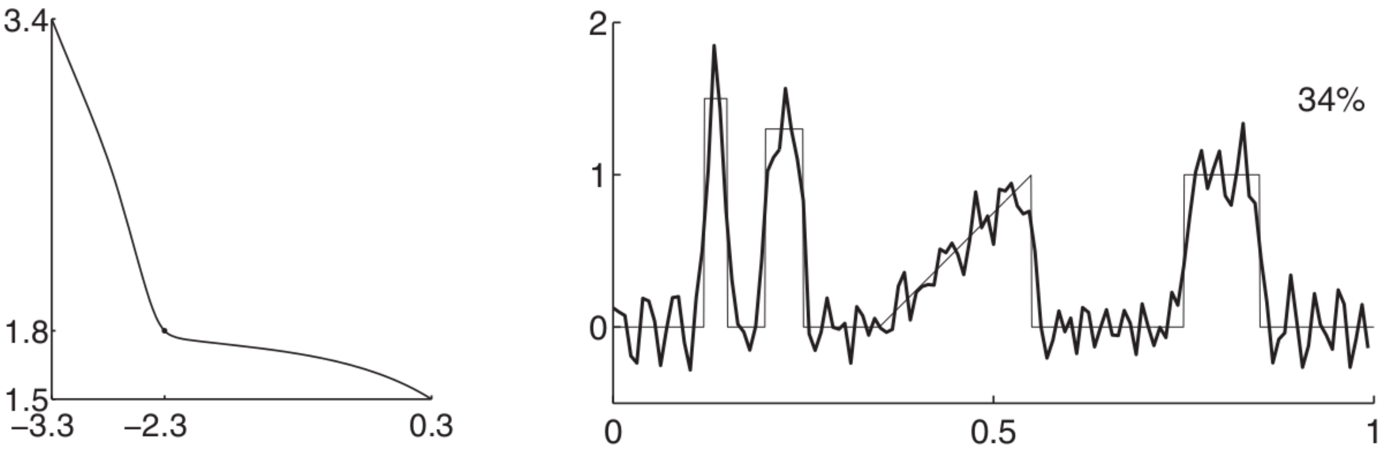


Figure 5.9. L-curve for the one-dimensional deconvolution problem.

Total variation regularization

Why? ->

- Edge preserving

Total variation regularization

Why? ->

- Edge preserving

What? ->

- Replace 2-norm by 1-norm in penality term of the generalized Tikhonov regularization

$$||A\mathbf{z} - \mathbf{m}||^2 + \alpha \sum_{j=1}^{n} |(L\mathbf{z})_j|$$

TV regularization

TV Definition: f is function defined on the interval $[a,b]$. TV is than:

$$TV(f) = \sup \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \tag{12}$$

where the supremum is over all partitions $a = x_0 < x_1 < ... < x_k = b$ of $[a,b]$

TV regularization

TV Definition: f is function defined on the interval $[a, b]$. TV is than:

$$TV(f) = \sup \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \tag{12}$$

where the supremum is over all partitions $a = x_0 < x_1 < ... < x_k = b$ of $[a, b]$

If differentiable (generalized also to higher dimensions)->

$$TV(f) = \int_{\Omega} |\Delta f(x)| dx \tag{13}$$

TV regularization: Edge preserving

Solution is blocker because sharp jumps are not strongly penalized.

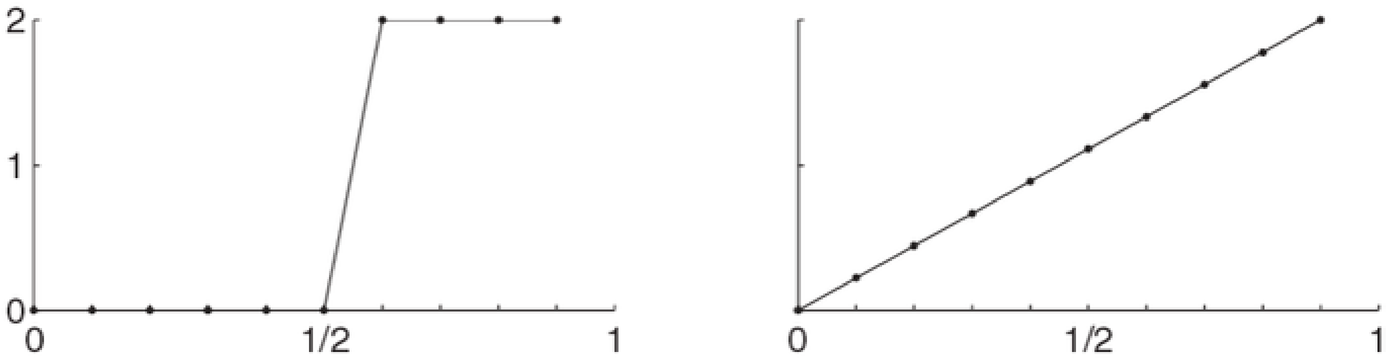


Figure 6.1. Two functions with interesting differences in their 1- and 2-norms. Left: **h**. Right: **f**.

$$||L\mathbf{f}||_2^2 = 44.44 \qquad ||L\mathbf{h}||_2^2 = 400$$

$$||L\mathbf{f}||_1 = 20 \qquad ||L\mathbf{h}||_1 = 20$$

TV regularization: Computation

- Medium-scale constrained quadratic programming
- Large-scale gradient-based minimization methods
- And other methods e.g. lagged diffusivity method; Lagrange multiplier methods; frame-based thresholding methods....

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

$$L\mathbf{f} =: \mathbf{v}_+ - \mathbf{v}_- \quad \text{with} \quad \mathbf{v}_\pm \in \mathbb{R}_+^{\ltimes} \tag{14}$$

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

$$Lf =: \mathbf{v}_+ - \mathbf{v}_- \quad \text{with} \quad \mathbf{v}_\pm \in \mathbb{R}_+^{\ltimes} \tag{14}$$

$$\text{old:} \quad ||A\mathbf{z} - \mathbf{m}||^2 + \alpha \sum_{j=1}^{n} |(L\mathbf{z})_j| \quad \text{new:} \quad ||A\mathbf{f}||^2 - 2\mathbf{m}^T A\mathbf{f} + \alpha \mathbf{1}^T \mathbf{v}_+ + \alpha \mathbf{1}^T \mathbf{v}_- \tag{15}$$

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

$$L\mathbf{f} =: \mathbf{v}_+ - \mathbf{v}_- \quad \text{with} \quad \mathbf{v}_\pm \in \mathbb{R}_+^{\ltimes} \tag{14}$$

$$\text{old:} \quad ||A\mathbf{z} - \mathbf{m}||^2 + \alpha \sum_{j=1}^{n} |(L\mathbf{z})_j| \quad \text{new:} \quad ||A\mathbf{f}||^2 - 2\mathbf{m}^T A\mathbf{f} + \alpha \mathbf{1}^T \mathbf{v}_+ + \alpha \mathbf{1}^T \mathbf{v}_- \tag{15}$$

because $||A\mathbf{f}||^2 = \mathbf{f}^T A^T A\mathbf{f}$

$$H := \begin{bmatrix} 2A^T A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \mathbf{h} = \begin{bmatrix} -2A^T\mathbf{m} \\ \alpha\mathbf{1} \\ \alpha\mathbf{1} \end{bmatrix} \quad y := \begin{bmatrix} f \\ \mathbf{v}_+ \\ \mathbf{v}_- \end{bmatrix} \tag{16}$$

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

   ...leading to

$$\arg\min_{\mathbf{y}} \frac{1}{2}\mathbf{y}^T H\mathbf{y} + \mathbf{h}^T\mathbf{y} \tag{17}$$

   with the constrains

$$L\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{2n} \end{bmatrix} - \begin{bmatrix} y_{2n+1} \\ \vdots \\ y_{3n} \end{bmatrix}$$

2. Solve

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

   ...leading to

   $$\arg\min_{\mathbf{y}} \frac{1}{2}\mathbf{y}^T H \mathbf{y} + \mathbf{h}^T \mathbf{y} \tag{17}$$

   with the constrains

   $$L \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{2n} \end{bmatrix} - \begin{bmatrix} y_{2n+1} \\ \vdots \\ y_{3n} \end{bmatrix}$$

2. Solve
   - the converted problem has 3n degrees of freedom, whereas the original has only n.

TV regularization: Computation with quadratic programming

1. Convert the problem to the standard form for quadratic programming.

   ...leading to

$$\arg\min_{\mathbf{y}} \frac{1}{2}\mathbf{y}^T H\mathbf{y} + \mathbf{h}^T\mathbf{y} \tag{17}$$

   with the constrains

$$L\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{2n} \end{bmatrix} - \begin{bmatrix} y_{2n+1} \\ \vdots \\ y_{3n} \end{bmatrix}$$

2. Solve
   - the converted problem has 3n degrees of freedom, whereas the original has only n.
   - in the two dimensional case there are 5n.

Result

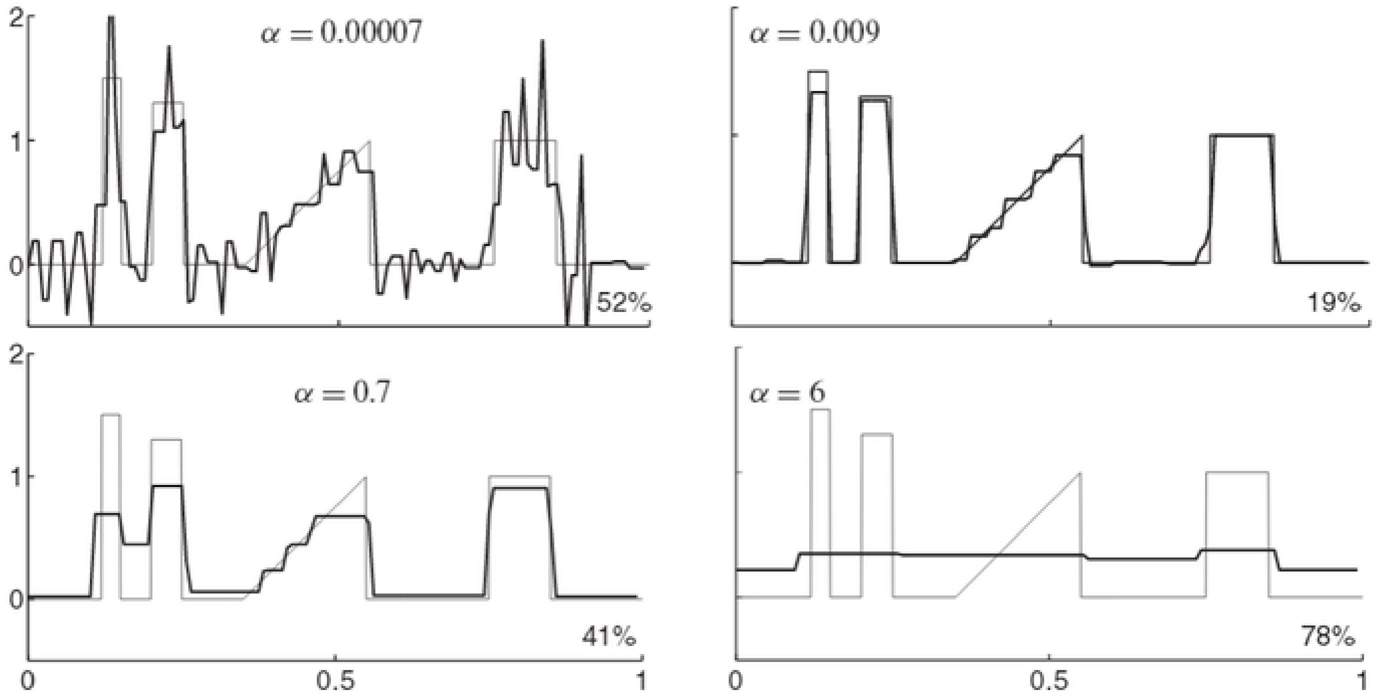


Figure 6.2. Total variation regularized reconstructions. The percentages shown are relative errors of reconstructions. Note the staircasing effect in the linear ramp part of the signal; this is a typical artefact of total variation inversion. Here $n = 128$.

Parameter choice: S-curve method

Priori information: number of nonzero coefficients in the true signal.

Compute for multiple parameters and choose the one with similar number.

Parameter choice: S-curve method

Priori information: number of nonzero coefficients in the true signal.

Compute for multiple parameters and choose the one with similar number.

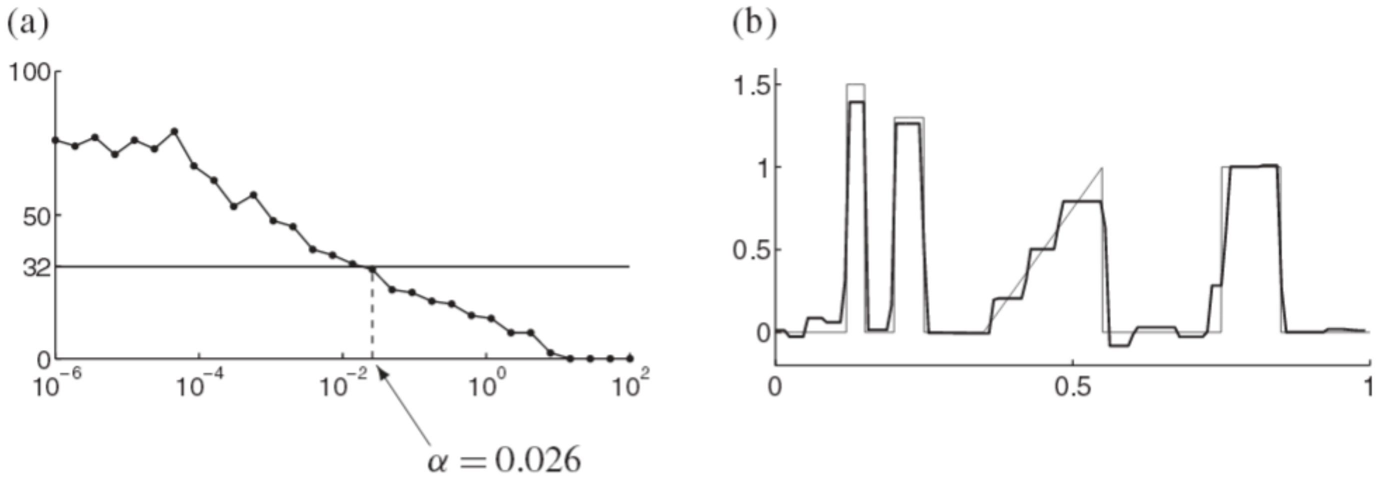


Figure 6.4. Sparsity-based choice of regularization parameter for total variation regularization. Here $n = 128$. (a) Number of jumps in the reconstruction as function of regularization parameter α. Note the logarithmic scale in the horizontal α-axis. (b) Reconstruction corresponding to the choice $\alpha = 0.026$ (thick line) and original signal (thin line).

TV regularization: Computation

- Medium-scale constrained quadratic programming
- **Large-scale gradient-based minimization methods**
- And other methods e.g. lagged diffusivity method; Lagrange multiplier methods; frame-based thresholding methods....

Gradient descent minimization method of Barzilai and Borwein

$$\arg\min_{\mathbf{f}} ||A\mathbf{f} - \mathbf{m}||_2^2 + \alpha |Lf|_1 = \arg\min_{\mathbf{f}} ||A\mathbf{f} - \mathbf{m}||_2^2 + \alpha \sum_{i=1}^{k} |f_i - f_{i-1}|$$

Gradient descent minimization method of Barzilai and Borwein

$$\arg\min_{\mathbf{f}} ||A\mathbf{f}-\mathbf{m}||_2^2+\alpha|Lf|_1 = \arg\min_{\mathbf{f}} ||A\mathbf{f}-\mathbf{m}||_2^2+\alpha\sum_{i=1}^{k}|f_i - f_{i-1}|$$

Approximate the L^1 norm with the $L^{1+\varepsilon}$:

$$|t|_\varepsilon = \sqrt{t^2+\varepsilon} \quad or \quad |t|_\varepsilon = \frac{1}{\varepsilon} log(cosh(\varepsilon t)) \tag{18}$$

Overview

Shortcoming of TV

- Loss of fine structures and contrast
- May lead to staircasing

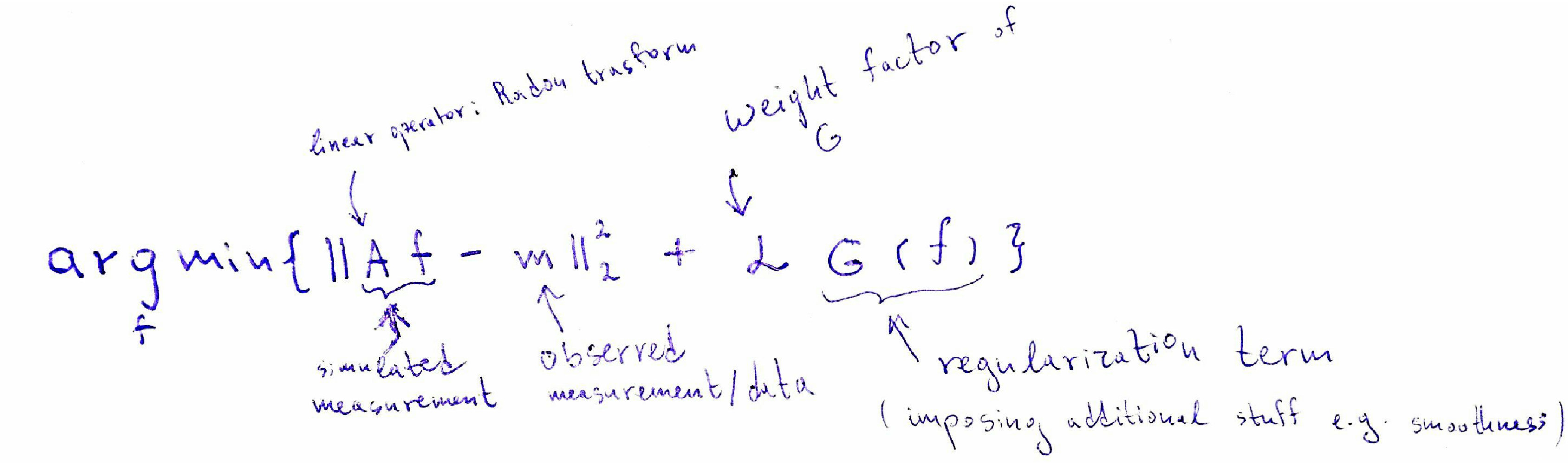
Shortcoming of TV

- Loss of fine structures and contrast
- May lead to staircasing

...leading to series expansion frameworks for reconstruction with sparsifying and edge-preserving dictionaries like shearlets and curvelets (for the regularization term)

Problem review

$$\arg\min_{\mathbf{f}} ||A\mathbf{f} - \mathbf{m}||^2 + \alpha G(f)$$



Sparse regularization

$$\arg\min_{\mathbf{f}} ||A\mathbf{f} - \mathbf{m}||^2 + \alpha G(f) \quad (19)$$

$G : f \longmapsto \| T f \|_1$ ← favors sparse solutions

↑ sparsifying operator

- T is a gradient operator (L) -> TV
- T - series expansion framework, e.g. with curvelets -> Curvelet Sparse Regularization

Sparse regularization

Sparse - continuous signal can be represented by a finite number of coefficients in a suitable basis

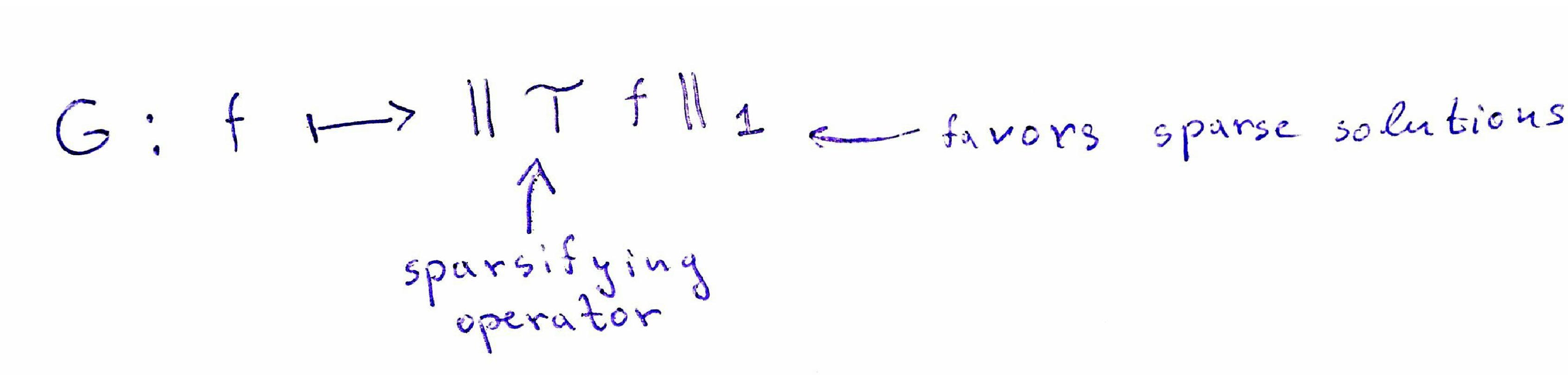
Sparse Regularization - Tf is supposed to contain relatively few nonzero values

Curvelet Sparse Regularization

Basis: curvelets

$T := C$ the curvelet transform

Minimizing algorithm



Problem: L^1-norm is *not* continuously differentiable => gradient descent does not work. Options:

1. Approximate L^1 with $L^{1+\varepsilon}$ (method of Barzilai Borwein with TV)
2. Splitting techniques like *Alternating Direction Method of Multipliers* (ADMM)

Curvelet Sparse Regularization

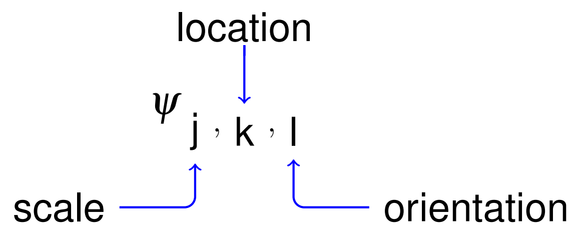
$T := C$ the curvelet transform

So lets look at L^1 with curvelets for the penality and ADMM for the minimization.

T := C the curvelet transform

CSR: the curvelet frame

Curvelets: family of functions



CSR: the curvelet frame

Curvelets: family of functions

location

$$\psi_{j,k,l}$$

scale orientation

So now we can expand any function $f \in L^2(\mathbb{R}^2)$:

$$f = \sum_{j,l,k} \langle \psi_{j,k,l}, f \rangle_{L^2} \psi_{j,k,l}$$

Minimizing algorithm: ADMM

Convert $\arg\min_f ||A\mathbf{f} - \mathbf{m}||^2 + \alpha|Cf|$

to

$$\arg\min_f ||A\mathbf{f} - \mathbf{m}||^2 + \alpha|c| \quad \text{s.t.} \quad Cf = c$$

Minimizing algorithm: ADMM

Convert $\arg\min_f ||A\mathbf{f} - \mathbf{m}||^2 + \alpha|Cf|$

to

$$\arg\min_f ||A\mathbf{f} - \mathbf{m}||^2 + \alpha|c| \quad \text{s.t.} \quad Cf = c$$

and after some fancy stuff (keyword: Lagragian) we arrive at ...

1. A linear inverse problem -> solve approximately with a gradiant method
2. Threshholding step
3. A simple dual update

Minimizing algorithm: ADMM

1. A linear inverse problem -> solve approximately with a gradient method
2. Thresholding step:
   S is the proximity operator to the L^1-norm, "soft-thresholding". Here the threshold being $\frac{\alpha}{\beta}$
3. A simple dual update

1. $(A^T A + \beta C^T C)(f^{k+1}) = (A^T m + \beta C^T (c^k + u^k))$
2. $c^{k+1} = S(C(f^{k+1}) + u^k)$ with $S(x) = \begin{cases} x - sgn(x)\frac{\alpha}{\beta} & |x| \geq \frac{\alpha}{\beta} \\ 0 & else \end{cases}$
3. $u^{k+1} = u^k + C(x^{k+1}) - z^{k+1}$

Parameter discussion

1. $(A^T A + \beta C^T C)(f^{k+1}) = (A^T m + \beta C^T(c^k + u^k))$
2. $c^{k+1} = S(C(f^{k+1}) + u^k)$ with $S(x) = \begin{cases} x - sgn(x)\frac{\alpha}{\beta} & |x| \geq \frac{\alpha}{\beta} \\ 0 & else \end{cases}$
3. $u^{k+1} = u^k + C(x^{k+1}) - z^{k+1}$

- α - our regular regularization parameter

Parameter discussion

1. $(A^T A + \beta C^T C)(f^{k+1}) = (A^T m + \beta C^T (c^k + u^k))$
2. $c^{k+1} = S(C(f^{k+1}) + u^k)$ with $S(x) = \begin{cases} x - sgn(x)\frac{\alpha}{\beta} & |x| \geq \frac{\alpha}{\beta} \\ 0 & else \end{cases}$
3. $u^{k+1} = u^k + C(x^{k+1}) - z^{k+1}$

- α - our regular regularization parameter
- β - coupling parameter

Parameter discussion

We could try some combinations out and choose the best one.

Parameter discussion

We could try some combinations out and choose the best one.

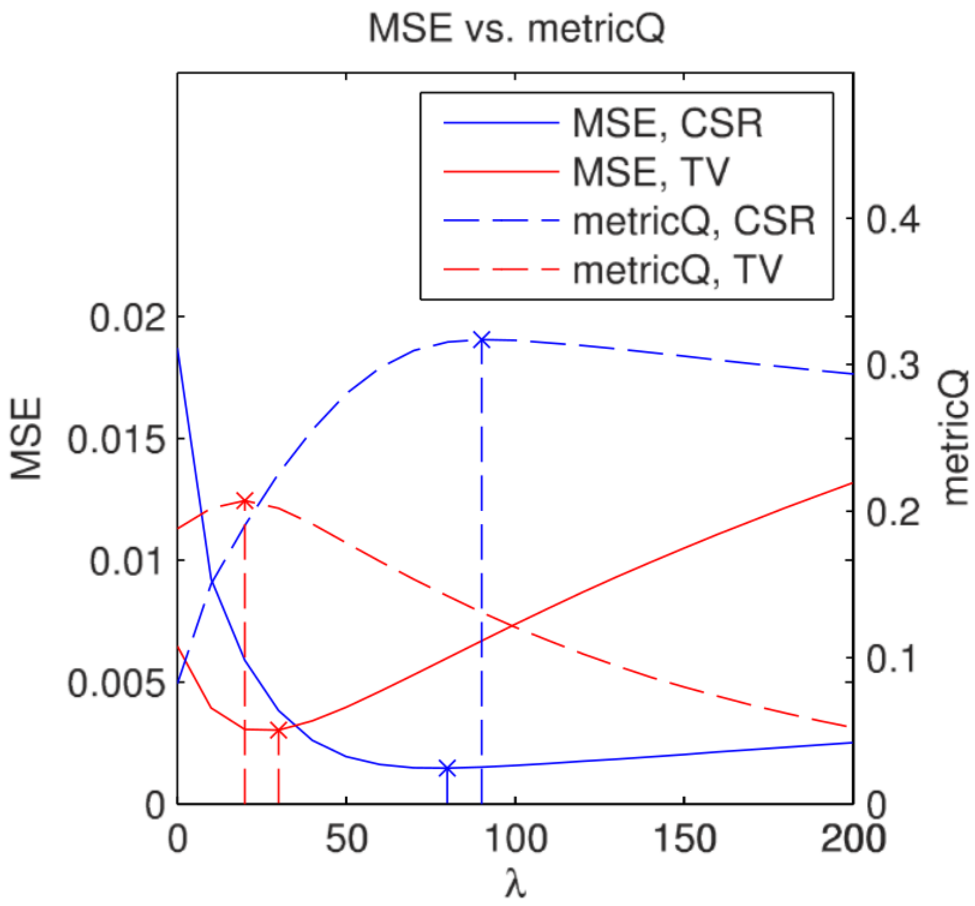
How do we know which is ”best”?

Parameter discussion

We could try some combinations out and choose the best one.
How do we know which is ”best”?

We employ a metric to judge the quality of reconstruction.

Parameter discussion



Parameter discussion

- Computational expense?
- Can we take advantage that ADMM is iterative?

CSR vs TV

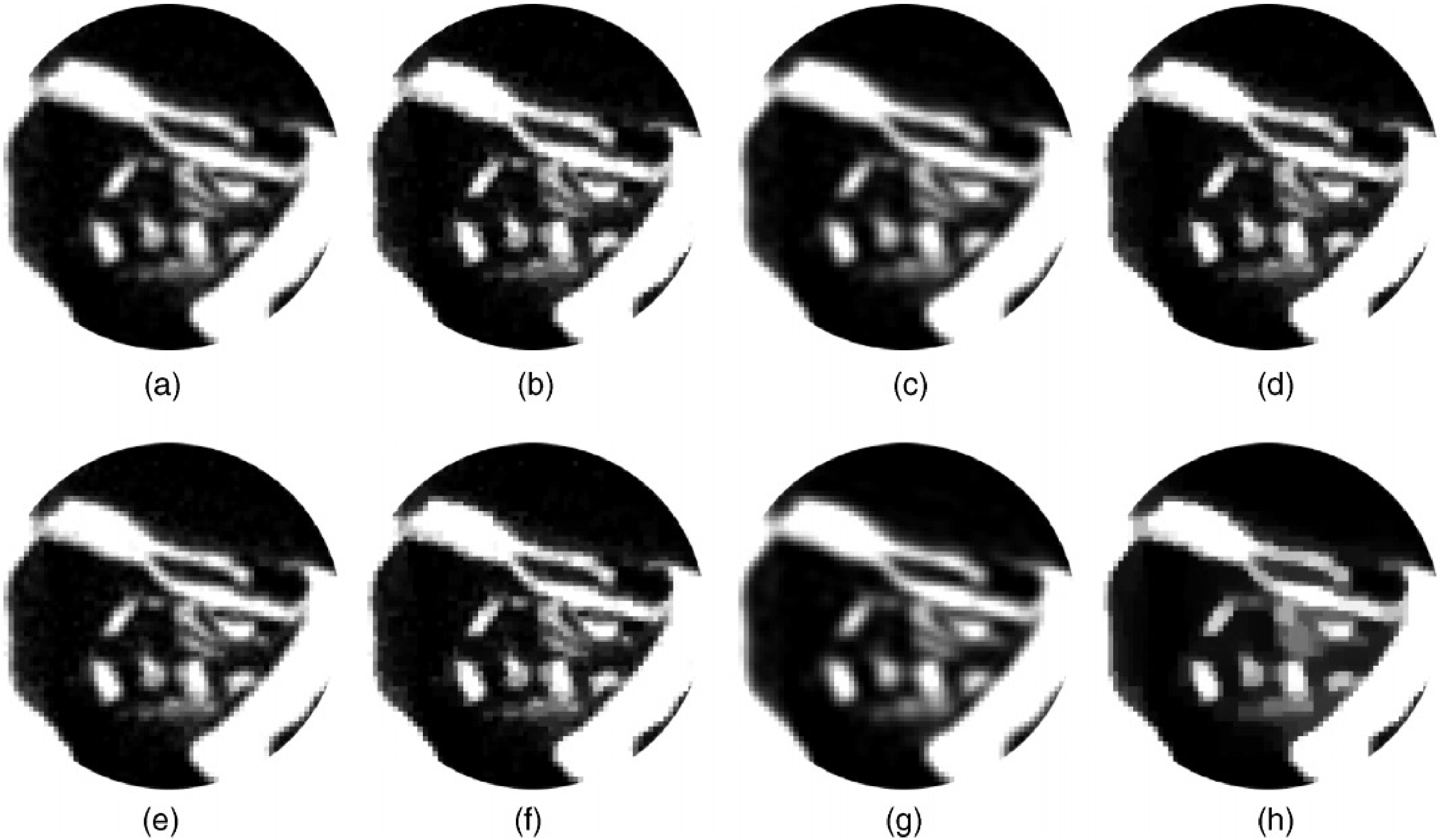
- CSR > TV structured regions; highly directional, high contrast features with smooth contrast variations
- CSR < TV homogeneous regions; CSR oscillating artifacts

CSR vs TV

- CSR $>$ TV structured regions; highly directional, high contrast features with smooth contrast variations
- CSR $<$ TV homogeneous regions; CSR oscillating artifacts

Let's see now some real stuff (stuff being a femur μCT)

CSR vs TV



(a) (b) (c) (d) (e) (f) (g) (h)

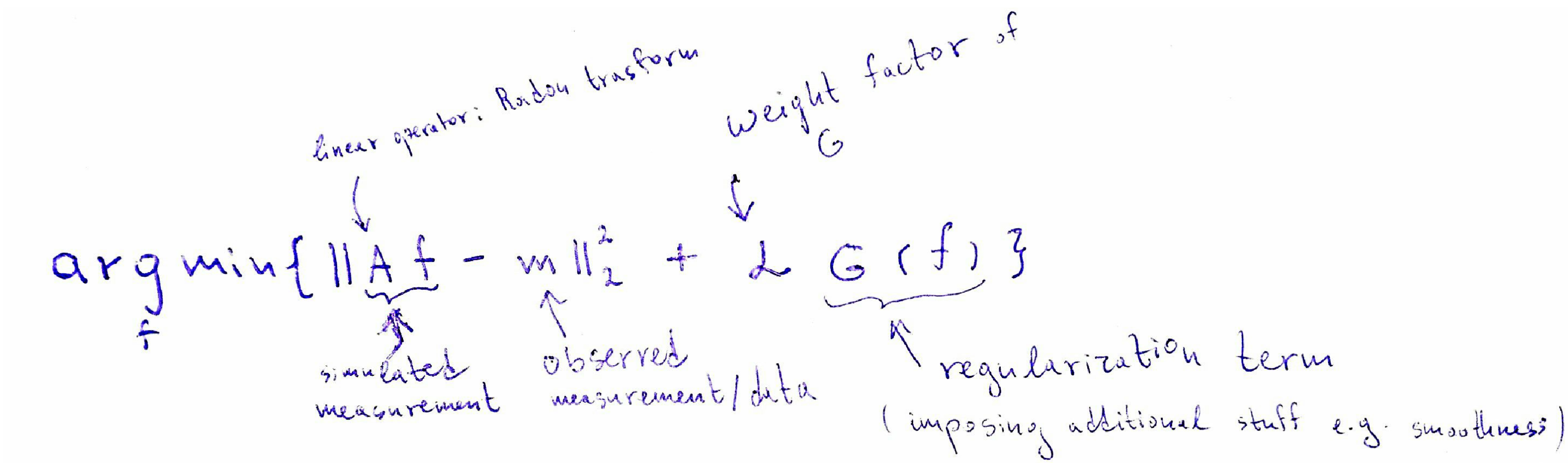
Summary

We must regularize!

We have to make choices about:

- The penality term.
- The parameters' choice.
- The computational approach.

Summary



- Tikhonov reg. favors smooth solutions & uses L^2-norm
- TV & CSR preserve edges & use L^1-norm
  - CSR > TV smooth image with edges along smooth curves
  - CSR < TV homogeneous regions

How do we choose?

We take our knowledge about the expected images and choose the best suitable method for it.

Basics of X-ray CT: Noise

Viktoria Markova

Sarntal, some September 2018

(a) 50 × 50 phantom

(b) Naïve inversion, ideal data, inverse crime

(c) Naïve inversion, data with 0.1% noise