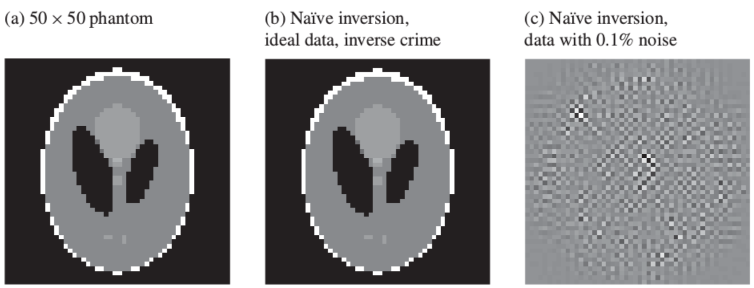
Basics of X-ray CT: Noise

Viktoria Markova

Sarntal, some September 2018



The problem

The Problem: Find f

find me

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

noise

measurement

linear operator

The problem

The Problem: Find f

find me

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

noise

measurement

linear operator: the Radon transform

The problem

Naive reconstruction:

$$f \quad A^{1}\mathbf{m} \tag{1}$$

The problem

Naive reconstruction: *”inverse crime”*

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

(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

$$v = \arg\min_{z \in \mathbb{R}^n} ||Az - m||^2 + a||z||^2 \qquad (2)$$

- Small residual $Av - m$
- v small in L^2 norm (prevents over tting)

Generalized Tikhonov regularization (priori information)

f is close to f :

$$v = \arg\min jj\,Az \quad m jj^2 + a\, jj\, z \quad f \; jj^2 \tag{3}$$

f is known to be smooth:

$$v = \arg\min jj\,Az \quad m jj^2 + a\, jj\, Lz jj^2 \tag{4}$$

or

$$v = \arg\min jj\,Az \quad m jj^2 + a\, jj\, L(z \quad f \;) jj^2 \tag{5}$$

L is a discretized differential operator/matrix

Tikhonov regularization: Computation

Stacked form of the non-generalized equation:

$$\begin{pmatrix} A \\ \sqrt{\alpha}\, I \end{pmatrix} f = \begin{pmatrix} m \\ 0 \end{pmatrix} \qquad (6)$$

written as

$$\tilde{A} f = \tilde{m} \qquad (7)$$

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

$$f = \tilde{A} \backslash \tilde{m} \qquad (8)$$

Generalized form:

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

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?

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 a, such that

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

If

$$||Pm|| <= \text{noise} <= ||m|| \tag{11}$$

then a 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 a
2. Compute the result v for each.
3. Plot $\log(\|Av - m\|); \log(\|Lv\|)$ for results
4. Observe the ”L curve graph”. ”Optimal” solution at the corner.

L-curve method: Example

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

$$\|Az - m\|^2 + \alpha \sum_{j=1}^{n} |(Lz)_j|$$

TV regularization

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

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

where the supremum is over all partitions $a = x_0 < x_1 < \dots < x_k = b$ 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 $a = x_0 < x_1 < \dots < x_k = b$ of $[a;b]$

If differentiable (generalized also to higher dimensions)->

$$TV(f) = \int |\nabla f(x)| dx \tag{13}$$

TV regularization: Edge preserving

Solution is blocker because sharp jumps are not strongly penalized.

$$\|Lf\|_2^2 = 44.44 \qquad \|Lh\|_2^2 = 400$$

$$\|Lf\|_1 = 20 \qquad \|Lh\|_1 = 20$$

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.

$$Lf =: v_+ \quad v \qquad \text{with} \quad v \quad 2\ R^n_+ \tag{14}$$

TV regularization: Computation with quadratic programming

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

$$Lf =: v_+ - v_- \quad \text{with} \quad v_\pm \in \mathbb{R}^n_+ \tag{14}$$

$$\text{old:} \quad ||Az - \mu||^2 + \alpha \sum_{j=1}^{n} |(Lz)_j| \quad \text{new:} \quad ||Af||^2 - 2\mu^T Af + \alpha 1^T v_+ + \alpha 1^T v_- \tag{15}$$

TV regularization: Computation with quadratic programming

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

$$Lf =: v_+ - v_- \quad \text{with} \quad v_\pm \in \mathbb{R}^n_+ \tag{14}$$

$$\text{old:} \quad \|Az - m\|^2 + \alpha \sum_{j=1}^{n} |(Lz)_j| \quad \text{new:} \quad \|Af\|^2 - 2m^T Af + \alpha 1^T v_+ + \alpha 1^T v_- \tag{15}$$

because $\|Af\|^2 = f^T A^T Af$

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

TV regularization: Computation with quadratic programming

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

   ...leading to

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

   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

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

   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

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

   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

Parameter choice: S-curve method

Priori information: number of nonzero coef cients 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 coef cients in the true signal.

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

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

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

Gradient descent minimization method of Barzilai and Borwein

$$\arg\min_f \|Af - m\|_2^2 + \alpha |Lf|_1 = \arg\min_f \|Af - m\|_2^2 + \alpha \sum_{i=1}^{k} |f_i - f_{i-1}|$$

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

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

Overview

Shortcoming of TV

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

Shortcoming of TV

- Loss of ne 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

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

Sparse regularization

$$\arg\min_{f} ||Af \quad m||^2 + a\,G(f) \tag{19}$$

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 nite number of coef cients 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+e} (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

location

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

scale — orientation

CSR: the curvelet frame

Curvelets: family of functions

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

(scale: j, location: k, orientation: l)

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 $\arg\min_f ||Af - m||^2 + a|Cf|$

to

$$\arg\min_f ||Af - m||^2 + a|c| \quad \text{s.t.} \quad Cf = c$$

Minimizing algorithm: ADMM

Convert $\arg\min_f \|Af - m\|^2 + \alpha |Cf|$

to

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

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 $\frac{a}{b}$
3. A simple dual update

1. $(A^T A + bC^T C)(f^{k+1}) = (A^T m + bC^T(c^k + u^k))$
2. $c^{k+1} = S(C(f^{k+1}) + u^k)$ with $S(x) = \begin{cases} x \quad \text{sgn}(x)\frac{a}{b} & |x| \quad \frac{a}{b} \\ 0 & \text{else} \end{cases}$
3. $u^{k+1} = u^k + C(x^{k+1}) \quad z^{k+1}$

Parameter discussion

1. $(A^T A + b C^T C)(f^{k+1}) = (A^T m + b C^T (c^k + u^k))$
2. $c^{k+1} = S(C(f^{k+1}) + u^k)$ with $S(x) = \begin{cases} x \quad sgn(x)\frac{a}{b} & jxj \quad \frac{a}{b} \\ 0 & \text{else} \end{cases}$
3. $u^{k+1} = u^k + C(x^{k+1}) \quad z^{k+1}$

a - our regular regularization parameter

Parameter discussion

1. $(A^T A + b C^T C)(f^{k+1}) = (A^T m + b C^T (c^k + u^k))$
2. $c^{k+1} = S(C(f^{k+1}) + u^k)$ with $S(x) = \begin{cases} x \quad \text{sgn}(x)\frac{a}{b} & jxj \quad \frac{a}{b} \\ 0 & \text{else} \end{cases}$
3. $u^{k+1} = u^k + C(x^{k+1}) \quad z^{k+1}$

a - our regular regularization parameter

b - 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

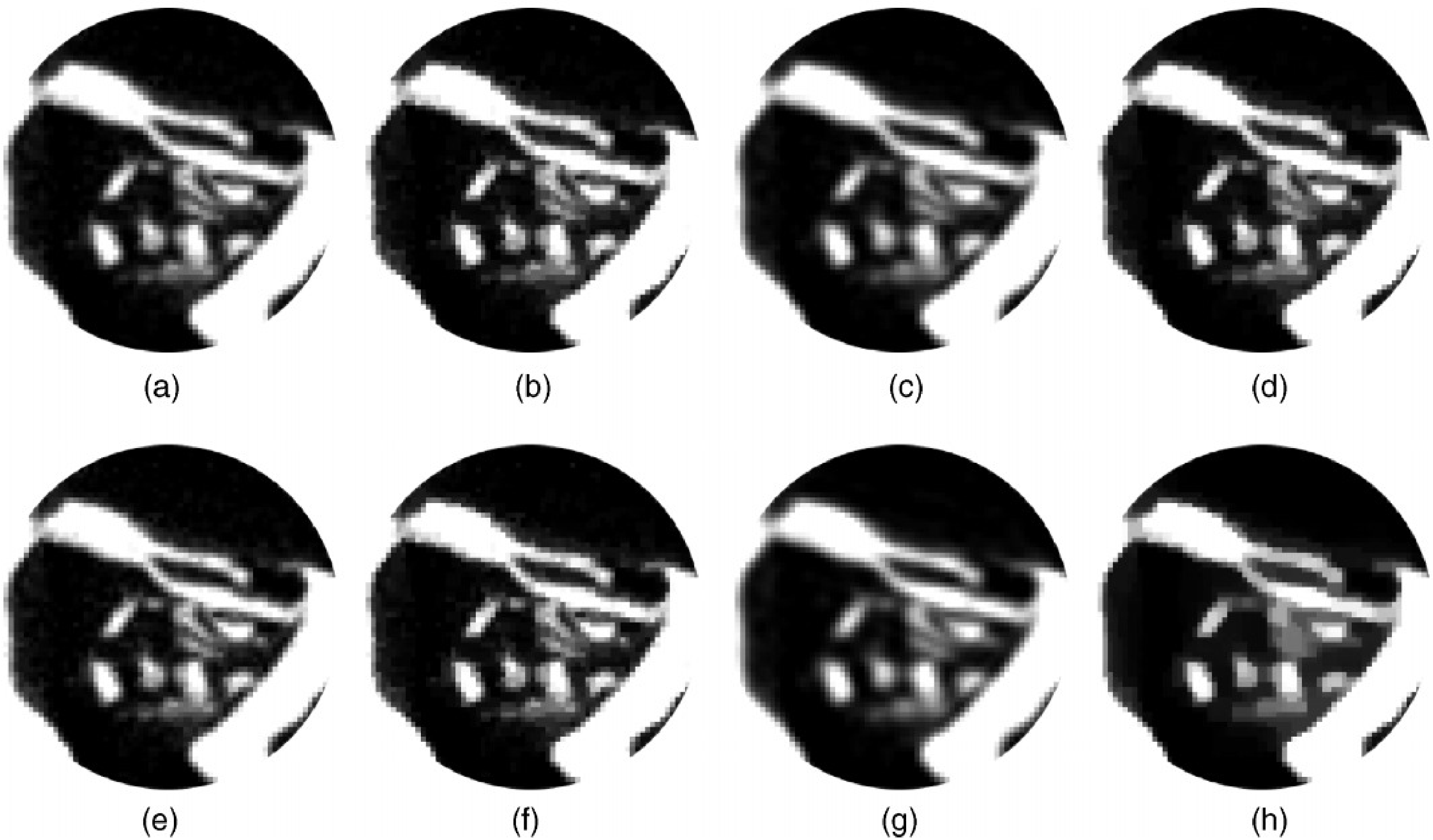
CSR vs TV

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 mCT)

CSR vs TV



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