Optimisation Methods for Tomography

Hugues Talbot

Université Paris-Est – ESIEE

July 8, 2016

Outline of the lecture

Section 1

Principles of tomography

X-ray absorption through a medium

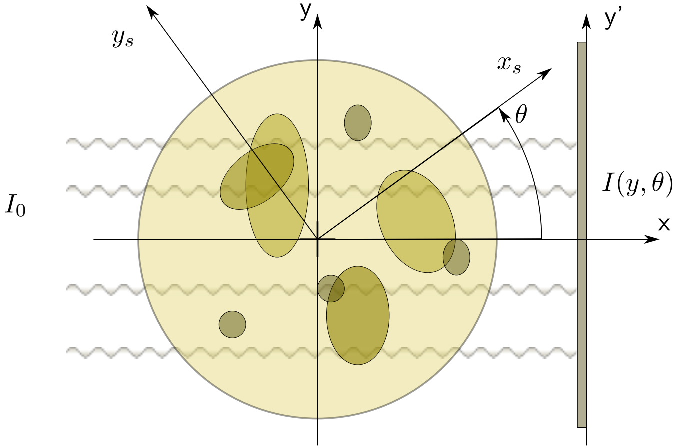


Figure: Illustrated principle of tomography.

Physical principle of tomography

- We assume the material under study absorbs some X-ray energy causing a loss of intensity.
- We assume that this loss is proportional to the intensity of the beam and the local absorption coefficient $\alpha(x, y)$. We write

$$\frac{\mathrm{d}I(x,y)}{\mathrm{d}x} = -\alpha(x,y)I(x,y), \tag{1}$$

- Solving this yields

$$I(x,y) = \exp(K)\exp(-\int \alpha(x,y)\mathrm{d}x) \tag{2}$$

- Boundary conditions are $K = I(-r, y) = I_0$, the intensity of the beam, and $\alpha(y, \theta) = \int_{-r}^{+r} \alpha(x_s, y_s)\, \mathrm{d}x$. With this

$$I(y,\theta) = I_0 \exp(-\alpha(y,\theta)) \tag{3}$$

Linear projection operator

- In a rotating frame, we write

$$\begin{aligned} x_s &= y\cos\theta + x\cos(\pi-\theta) \\ y_s &= y\sin\theta + x\sin(\pi-\theta), \end{aligned} \tag{4}$$

which is the parametric equation of the X-ray beam with ordinate y.

- We define $S(y,\theta) = -\log(I(y,\theta)/I_0)$. This is a linear function of (x_s, y_s)
- We denote this operator Θ, the tomography projection operator:

$$\begin{aligned} \Theta :& \mathbb{R}^2 \longmapsto \mathbb{R} \times [0, 2\pi[\\ & \alpha(x_s, y_s) \longrightarrow S(y,\theta) \end{aligned} \tag{5}$$

- $S(y,\theta)$ is the ideal observed data: the *sinogram*.

Direct problem

- Calculating the sinogram from an image is easy.
- In discrete form, with a fixed number of angles $n\theta$ and I is a $nx \times ny$ image, the Θ operator can be modelled by a matrix $\mathbf{T}$ of dimensions $[nx \times ny, ny \times n_\theta]$
- Computing the sinogram consists of a matrix multiplication

$$\mathsf{S} = \mathbf{T}\mathsf{I} \tag{6}$$

- This is an example of a *direct* problem.

(a) (b)

Tomography operator

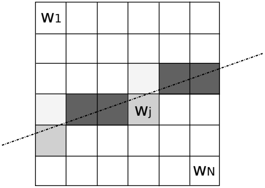


Figure: Determining the tomography operator

- Determining the tomography operator is not so trivial however.

$$S(y,\theta) = \sum_{i \in I} w_{i,\theta,y} I(x,y) \tag{7}$$

- Various choices of w exist for parallel beam, cone beam, 2D, 3D, etc.
- Analytical expression may be difficult to find, however always solvable in practice.

Inverse problem

- However, going the other way is hard: an *inverse* problem.
- The main problems are due to noise and limited number of projections. In reality we observe

$$\mathsf{S} = \mathbf{T}\mathsf{I} + \varepsilon, \tag{8}$$

  where ε is some noise, often modelled by a white additive Gaussian noise. More complex models exist (Poisson-Gauss, Rician, exponential, etc).
- The adjoint operator

$$\mathsf{J} = \mathbf{T}^\top \mathsf{S} \tag{9}$$

  is the *back-projection* operator.
- $\mathbf{T}$ is not square, not invertible, and usually *ill-conditioned*.
- The minimum norm solution

$$\mathsf{I}^\star = \mathbf{T}^\top (\mathbf{T}\mathbf{T}^\top)^{-1} \mathsf{S} \tag{10}$$

  is the *filtered back-projection* operator.

Tomography with limited angles

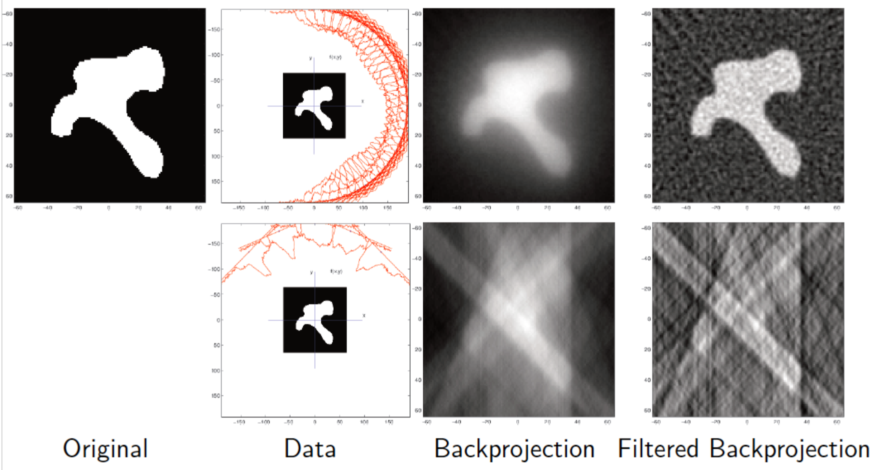
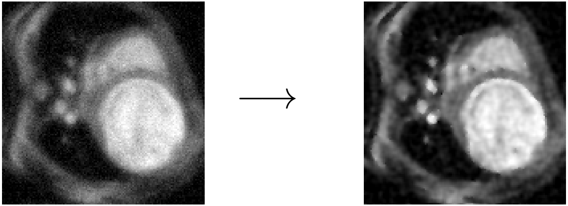


Figure: Tomography with limited angular resolution

Section 2

Inverse problems in imaging

Motivation: inverse problems in imaging



- Images we observe are nearly always blurred, noisy, projected versions of some “reality”.
- We wish to dispel the fog of acquisition by removing all the artefacts as much as possible to observe the “real” data.
- This is an *inverse* problem.

Maximum Likelihood

- We want to estimate some statistical parameter θ on the basis of some observation x. If f is the sampling distribution, $f(x|\theta)$ is the probability of x when the population parameter is θ. The function

$$\theta \mapsto f(x|\theta)$$

is the *likelihood*. The Maximum Likelihood estimate is

$$\hat{\theta}_{\mathsf{ML}}(x) = \underset{\theta}{\operatorname{argmax}}\ f(x|\theta)$$

- E.g, if we have a linear operator $\mathbf{H}$ (in matrix form) and Gaussian deviates, then

$$\underset{\mathsf{x}}{\operatorname{argmax}}\ f(x) = -\|\mathbf{H}\mathsf{x} - \mathsf{y}\|_2^2 = -\mathsf{x}^\top \mathbf{H}^\top \mathbf{H}\mathsf{x} + 2\mathsf{y}^\top H x - y^\top y$$

is a quadratic form with a unique maximum, provided by

$$\nabla f(\mathsf{x}) = -2\mathbf{H}^\top \mathbf{H} x + 2\mathbf{H}^\top \mathsf{y} = 0 \rightarrow \theta = (\mathbf{H}^\top \mathbf{H})^{-1}\mathbf{H}^\top \mathsf{y}$$

Strengths and drawbacks of MLE

- When possible, MLE is fast and effective. Many imaging operators have a MLE interpretation:
  - Gaussian smoothing ;
  - Wiener filtering ;
  - Filtered back projection for tomography ;
  - Principal component analysis ...
- However these require a very descriptive model (with few degrees of freedom) and a lot of data, typically unsuitable for images because we do not have a suitable model for natural images.
- When we do not have all these hypotheses, sometimes the Bayesian Maximum A Posteriori approach can be used instead.

Maximum A Posteriori

- If we assume that we know a *prior* distribution g over θ, i.e. some *a-priori* information. Following Bayesian statistics, we can treat θ as a random variable and compute the *posterior* distribution of θ:

$$\theta \mapsto f(\theta|x) = \frac{f(x|\theta)g(\theta)}{\int_{\vartheta \in \Theta} f(x|\vartheta)g(\vartheta)d\vartheta}$$

  (i.e. the Bayes theorem).

- Then the Maximum a Posteriori is the estimate

$$\hat{\theta}_{MAP}(x) = \underset{\theta}{\operatorname{argmax}}\ f(\theta|x) = \underset{\theta}{\operatorname{argmax}}\ f(x|\theta)g(\theta)$$

- MAP is a *regularization* of ML.

Section 3

Concepts in optimisation

Introduction

- Mathematical optimization is a domain of applied mathematics relevant to many areas including statistics, mechanics, signal and image processing.
- Generalizes many well known techniques such as least squares, linear programming, convex programming, integer programming, combinatorial optimization and others.
- In this talk we will overview both the continuous and discrete formulations.
- We follow the notations of Boyd & Vandeberghe [2].

General form

Cost function and constraints

An optimization problem generally has the following form

$$\begin{aligned} \text{minimize } & f_0(x) \\ \text{subject to } & f_i(x) \leq b_i, i = 1, \ldots, m \end{aligned} \tag{11}$$

$x = (x_1, \ldots, x_n)$ is a vector of $\mathbb{R}^n$ called the *optimization variable* of the problem; $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the *cost function* functional; the $f_i : \mathbb{R}^n \to \mathbb{R}$ are the *constraints* and the b_i are the *bounds* (or limits).
A vector $x^\star$ is is *optimal*, or is a solution to the problem, if it has the smallest objective value among all vectors that satisfy the constraints.

Types of optimization problems

- The type of the variables, the cost function and the constraints determine the type of problems we are dealing with.
- Optimization problems, in their most general form, are usually unsolvable in practice. NP-complete problems (traveling salesperson, subset-sum, etc) can classically be put in this form and so can many NP-hard problems.
- Some mathematical regularity is necessary to be able to find a solution: for example, linearity or convexity in all the functions.
- Requiring integer solutions usually, but not always, makes things much harder: Diophantine vs linear equations for instance.

Resolution of optimisation problems

The resolution of an optimisation problem depends on its form. In order of complexity, we can solve optimisation problems:

- In closed form solution (some regression problems)
- If convex: by some iterative descent-like method, yielding a global optimum. Note: may work in the non-differentiable case.
- If non-convex, but regular in some other way (differentiable, quasi-convex, ...): iterative descent-like, converging to a local optimum (or a critical point).
- If combinatorial, usually NP-hard, some exceptions: transport problems (graph cuts, transshipment problems).
- If all else fails: brute force, meta-heuristics.

Example closed form: least-squares

Least squares with no constraints

$$\text{minimize } f_0(x) = \|\mathbf{H}\mathsf{x} - \mathsf{b}\|_2^2 = \sum_{i=1}^{k} a^\intercal x_i - b_i \qquad (12)$$

The system is quadratic, so convex and differentiable. The solution to (12) is unique and reduces to the linear equation

$$(\mathbf{H}^\intercal\mathbf{H})\mathsf{x} = \mathbf{H}^\intercal\mathsf{b}. \quad \text{(normal equation)} \qquad (13)$$

The analytical solution is $\mathsf{x} = (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal b$, however $\mathbf{H}^\intercal\mathbf{H}$ should never be calculated, much less the inverse, for numerical reasons.

Regularization: Tikhonov

Even with something as simple as least-squares, if $\mathbf{H}$ is ill-conditioned, the solution will be very sensitive to noise, e.g. in the example of deconvolution or tomography. One solution is to use regularization.

Ill-posed least-squares problems

The simplest regularization strategy is due to Tikhonov [8].

$$\text{minimize } f_0(x) = \|\mathbf{H}\mathsf{x} - \mathsf{b}\|_2^2 + \|\Gamma \mathsf{x}\|_2^2, \tag{14}$$

where Γ is a well-chosen operator, e.g. λI or ∇x or a wavelet operator. The solution is given analytically by

$$x = (\mathbf{H}^\intercal \mathbf{H} + \Gamma^\intercal \Gamma)^{-1} \mathbf{H}^\intercal \mathsf{b} \tag{15}$$

Example iterative: linear programming

Linear programming with constraints

$$\begin{aligned} &\text{minimize } c^\intercal x \\ &\text{subject to } a_i^\intercal x \leq b_i; i = 1, \ldots, n \end{aligned} \tag{16}$$

- No analytical solution.
- Well established family of algorithms: the Simplexe (Dantzig 1948) ; interior-point (Karmarkar 1984)
- Not always easy to recognize. Important for compressive sensing.

Continuous image restoration model

- We suppose there exists some unknown image $\overline{\mathrm{x}} \in \mathbb{R}^N$.
- However we do observe some data $\mathrm{y} \in \mathbb{R}^Q$ via some linear operator H, which is corrupted by some noise:

$$\mathrm{y} = \mathbf{H}\overline{\mathrm{x}} + \varepsilon, \qquad \mathbf{H} \in \mathbb{R}^{Q \times N}$$

$\overline{\boldsymbol{x}}$

$\boldsymbol{y}$

Recovery

- We seek to recover a good approximation $\hat{x}$ of $\bar{x}$ from $\mathbf{H}$ and y.
- H can be:
  - Model for camera, including defocus and motion blur
  - MRI, PET,
  - X-Ray tomography
  - ...
- ε often modeled by Additive White Gaussian Noise, but can be Poisson, Poisson Gauss, Rician, etc.

Simplest case: least squares:

$$\hat{x} = \mathrm{argmin}_x \|\mathbf{H}x - y\|_2^2$$

analytical, simple, effective, but not robust to outliers.

Recovery

- We seek to recover a good approximation $\hat{x}$ of $\bar{x}$ from $\mathbf{H}$ and y.
- H can be:
  - Model for camera, including defocus and motion blur
  - MRI, PET,
  - X-Ray tomography
  - ...
- ε often modeled by Additive White Gaussian Noise, but can be Poisson, Poisson Gauss, Rician, etc.

Tikhonov regularization:

$$\hat{x} = \mathrm{argmin}_x \|\Gamma x\|_2^2 + \lambda \|\mathbf{H}x - y\|_2^2$$

reflect the *prior* assumption that we want to avoid large x. Also analytical and more robust but not sparse.

Recovery

- We seek to recover a good approximation $\hat{x}$ of $\bar{x}$ from $\mathbf{H}$ and y.
- H can be:
  - Model for camera, including defocus and motion blur
  - MRI, PET,
  - X-Ray tomography
  - ...
- ε often modeled by Additive White Gaussian Noise, but can be Poisson, Poisson Gauss, Rician, etc.

Enforced sparsity:

$$\hat{x} = \text{argmin}_x \|\Gamma x\|_0 + \lambda \|\mathbf{H}x - y\|_2$$

If we know x to be sparse (many zero elements) in some space (e.g. Wavelets). Highly non-convex.

Recovery

- We seek to recover a good approximation $\hat{\mathrm{x}}$ of $\bar{\mathrm{x}}$ from $\mathbf{H}$ and y.
- H can be:
  - Model for camera, including defocus and motion blur
  - MRI, PET,
  - X-Ray tomography
  - ...
- ε often modeled by Additive White Gaussian Noise, but can be Poisson, Poisson Gauss, Rician, etc.

Compressive sensing:

$$\hat{\mathrm{x}} = \mathrm{argmin}_{\mathrm{x}} \|\Gamma \mathrm{x}\|_1 + \lambda \|\mathbf{H}\mathrm{x} - \mathrm{y}\|_2$$

If we know x to be sparse (many zero elements) in some space (e.g. Wavelets). Smallest convex approximation of the ℓ_0 pseudo-norm.

Formal context

Penalized optimization problem

Find

$$\min_{\mathsf{x}\in\mathbb{R}^N} \big(F(\mathsf{x}) = \Phi(\mathbf{H}\mathsf{x} - \mathsf{y}) + \lambda R(\mathsf{x})\big),$$

Φ ⇝ Fidelity to data term, related to noise

R ⇝ Regularization term, related to some *a priori* assumptions

λ ⇝ Regularization weight

Here, $\boldsymbol{x}$ is **sparse** in a dictionary $\mathcal{G}$ of analysis vectors in $\mathbb{R}^N$

$$F_0(\mathsf{x}) = \Phi(\mathbf{H}\mathsf{x} - \mathsf{y}) + \lambda\,\ell_0(\Gamma\mathsf{x})$$

Formal context

Penalized optimization problem

Find

$$\min_{x \in \mathbb{R}^N} \big(F(x) = \Phi(\mathbf{H}x - y) + \lambda R(x) \big),$$

Φ $\rightsquigarrow$ Fidelity to data term, related to noise

R $\rightsquigarrow$ Regularization term, related to some *a priori* assumptions

λ $\rightsquigarrow$ Regularization weight

Here, $\boldsymbol{x}$ is **sparse** in a dictionary $\mathcal{G}$ of analysis vectors in $\mathbb{R}^N$

$$F_\delta(x) = \Phi(\mathbf{H}x - y) + \lambda \sum_{c=1}^{C} \psi_\delta(\Gamma_c^\top x)$$

where ψ_δ is a differentiable, non-convex approximation of the ℓ_0 norm.

Non-convex optimization

- The current frontier.
- Many interesting applications thought to be very hard to solve: blind deblurring
- Many current methods extend to the Non-Convex case
- Generally only a local minimum is reached, but this might be OK. The miimum might be of high quality : stochastic optimization.
- For instance: see results achieved by deep-learning methods.

ℓ_2-ℓ_0 regularization functions

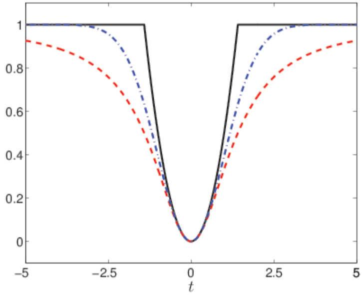
We consider the following class of potential functions:

1. $(\forall \delta \in (0, +\infty))$ ψ_δ is differentiable.
2. $(\forall \delta \in (0, +\infty))$ $\lim_{t\to\infty} \psi_\delta(t) = 1$.
3. $(\forall \delta \in (0, +\infty))$ $\psi_\delta(t) = \mathcal{O}(t^2)$ for small t.

Examples:

– – – $\psi_\delta(t) = \frac{t^2}{2\delta^2 + t^2}$

– · –· $\psi_\delta(t) = 1 - \exp(-\frac{t^2}{2\delta^2})$



Majorize-Minimize principle *[Hunter04]*

Objective: Find $\hat{\boldsymbol{x}} \in \operatorname{Arg\,min}_{\boldsymbol{x}} F_\delta(\boldsymbol{x})$

For all $\boldsymbol{x}'$, let $Q(., \boldsymbol{x}')$ a *tangent majorant* of F_δ at $\boldsymbol{x}'$ i.e.,

$$Q(\boldsymbol{x}, \boldsymbol{x}') \geqslant F_\delta(\boldsymbol{x}), \quad \forall \boldsymbol{x},$$
$$Q(\boldsymbol{x}', \boldsymbol{x}') = F_\delta(\boldsymbol{x}')$$

MM algorithm:

$$\forall j \in \{0, \ldots, J\},$$

$$\boldsymbol{x}^{j+1} \in \operatorname{Arg\,min}_{\boldsymbol{x}} Q(\boldsymbol{x}, \boldsymbol{x}^j)$$

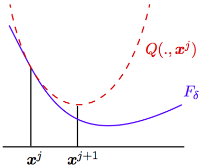
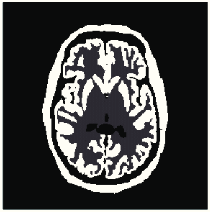
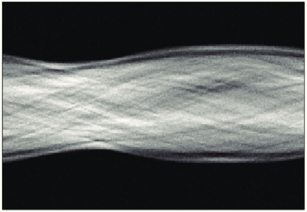


Image reconstruction



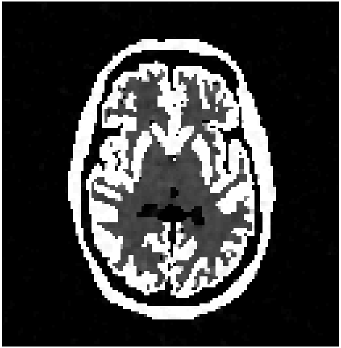
Original image $\overline{\boldsymbol{x}}$
128×128



Noisy sinogram $\boldsymbol{y}$
SNR=25 dB

- $\mathsf{y} = \mathbf{T}\overline{\mathsf{x}} + \varepsilon$ with $\begin{cases} \mathbf{T} & \text{Projection operator} \\ \varepsilon & \text{Gaussian noise} \end{cases}$
- $\hat{\mathsf{x}} \in \operatorname{Arg\,min}_{\mathsf{x}} \left(\frac{1}{2} \|\mathbf{T}\mathsf{x} - \mathsf{y}\|^2 + \lambda \sum_c \psi_\delta(\Gamma_c^\top \mathsf{x}) \right)$
- Non convex penalty / convex penalty

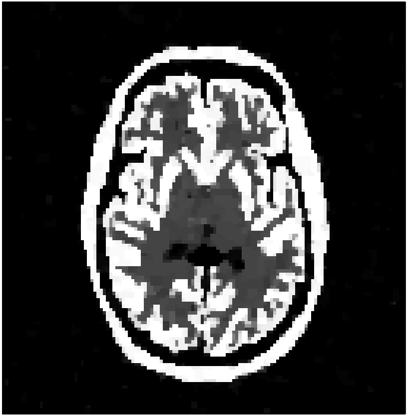
Results: Non convex penalty



Reconstructed image
SNR $= 20.4$ dB

MM-MG algorithm:
Convergence in 134 s

Results: Convex penalty



Reconstructed image
SNR $= 18.4$ dB

MM-MG algorithm:
Convergence in 60 s

Section 4

Conclusion

Conclusion

- Optimization is a very powerful, general methodology for solving inverse problems in variational form.
- We've drawn a panorama of interesting methodologies in image processing
- Generally optimization problems are unsolvable without some regularity assumptions. There exist a trade-off between the generality of a framework and the efficiency of associated algorithms.

Luigi Ambrosio and Vincenzo Maria Tortorelli.
Approximation of functional depending on jumps by elliptic functional via t-convergence.
Communications on Pure and Applied Mathematics, 43(8):999–1036, 1990.

S.P. Boyd and L. Vandenberghe.
Convex optimization.
Cambridge Univ Press, 2004.

P. L. Combettes and J.-C. Pesquet.
Proximal splitting methods in signal processing.
In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer-Verlag, New York, 2010.

P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot.
An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images.
Medical Imaging, IEEE Transactions on, 27(4):425–441, 2008.

L.J. Grady and J.R. Polimeni.

Discrete Calculus: Applied Analysis on Graphs for Computational Science.

Springer Publishing Company, Incorporated, 2010.

D. Mumford and J. Shah.
Optimal approximations by piecewise smooth functions and associated variational problems.
Communications on pure and applied mathematics, 42(5):577–685, 1989.

Gabriel Taubin, Tong Zhang, and Gene H. Golub.
Optimal surface smoothing as filter design.
In *Proceedings of the 4th European Conference on Computer Vision - Volume I*, ECCV '96, pages 283–292, London, UK, 1996. Springer-Verlag.

Andrey Nikolayevich Tikhonov.
On the stability of inverse problems.
In *Dokl. Akad. Nauk SSSR*, volume 39, pages 195–198, 1943.