Chapter 3
Variational Image Restoration

Variational Methods in Imaging

Feb 2021

Yvain QUÉAU
GREYC-CNRS
ENSICAEN - Université de Caen Normandie

Overview

Overview

1. **Inverse Problems and Variational Methods**
2. Variational Methods and Discrete Optimization
3. Variational Methods, Euler-Lagrange Equations and Diffusion
4. Variational Methods and Bayesian Inference
5. Numerical Solving
6. Conclusions

Inverse Problems, Ill-Posedness and Regularization

In mathematics, the conversion of measurement data into information about the observed object or the observed physical system is referred to as an inverse problem.

Following Hadamard (1902), a mathematical problem is called well-posed iff:

1. A solution exists.
2. The solution is unique.
3. The solution's behavior changes continuously with the initial conditions.

Inverse problems are often ill-posed. Since the measurement data is often not sufficient to uniquely characterize the observed object or system, one introduces prior knowledge to disambiguate which solutions are apriori more likely. In the context of variational methods this prior knowledge gives rise to the regularity term.

Variational Methods

Variational methods are a specific class of optimization methods. The key idea is to define cost functionals over a continuous solution space and to compute optima by solving the corresponding extremality principle.

Variational methods allow to solve respective problems in a mathematically transparent manner. Instead of performing a heuristic sequence of processing steps one starts by defining what properties a solution should have. Once these are fixed, the appropriate algorithm can be derived "automatically".

Variational methods are particularly suited for infinite-dimensional problems. They are among the top performing methods for:

- image denoising, deblurring, super-resolution
- image segmentation
- motion estimation
- dense 3D reconstruction
- tracking

Advantages of Variational Methods

Variational methods have many advantages over traditional multi-step approaches (such as the Canny edge detector):

- A mathematical analysis of the cost function allows statements regarding the existence, uniqueness and stability of solutions to a given problem.
- In traditional multistep processes the interplay of consecutive steps is often complex and intransparent. It is typically unclear how modifying or replacing one component affects the subsequent steps.
- Optimization methods are based on transparent and explicitly formulated assumptions, with no “hidden” assumptions.
- In general, optimization methods have fewer parameters. The meaning of each parameter is fairly obvious.
- Optimization methods are easily combined in a transparent manner (by adding respective cost functions).

A Simple Example: Image Denoising

Let $f : \Omega \to \mathbb{R}$ be an input greylevel image corrupted by noise. The goal is to compute a denoised version u of the image f.

The desired function u should fulfill two criteria:

- The function u should be similar to f.
- The function u should be spatially smooth.

Both criteria can be combined in the following cost function (or energy):

$$E(u) = E_{data}(u, f) + \lambda\, E_{smoothness}(u),$$

where the first term measures the similarity of u and f and the second term measures the smoothness of u. A weighting or regularization parameter $\lambda \geq 0$ specifies the relative importance of smoothness versus data fit.

Most variational approaches have the above form. They merely differ in how the similarity term (data term) and the smoothness term (regularizer) are defined.

Image Restoration: Denoising

Image restoration is a classical inverse problem: Given an observed image $f : \Omega \to \mathbb{R}$ and a (typically stochastic) model of an image degradation process, we want to restore the original image $u : \Omega \to \mathbb{R}$.

Image denoising is an example of image restoration where we assume that the true image u is corrupted by (additive) noise:

$$f = u + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma^2).$$

Rudin, Osher, Fatemi (1992) denoise f by minimizing a quadratic data term with Total Variation (TV) regularization:

$$\min_u \frac{1}{2} \int |u - f|^2 dx + \lambda \int |\nabla u| dx.$$

This gives rise to the necessary optimality condition (Euler-Lagrange equation)

$$u - f - \lambda \operatorname{div}\Big(\underbrace{\frac{1}{|\nabla u|}}_{\text{diffusivity !}} \nabla u \Big) = 0.$$

Other noise models and regularizers are conceivable.

Image Restoration: Denoising

original

noisy

denoised

(Goldlücke, Strekalovskiy, Cremers, SIAM J. Imaging Sci. ’12)

Image Restoration: Deblurring

A prototypical blur model is given by

$$f = A * u + \eta \qquad \eta \sim \mathcal{N}(0, \sigma^2),$$

with a blur kernel A.

In a variational setting, this process can be inverted by minimizing the TV deblurring functional:

$$\min_u \frac{1}{2} \int |A * u - f|^2 dx + \lambda \int |\nabla u| dx.$$

For symmetric kernels A, the optimality condition (Euler-Lagrange equation) is given by:

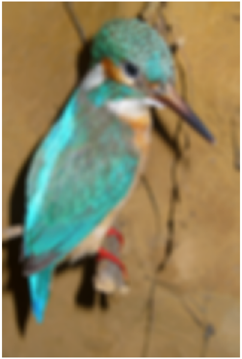
$$A * (A * u - f) - \lambda \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0,$$

and the gradient descent equation

$$\frac{\partial u}{\partial t} = -A * (A * u - f) + \lambda \, \mathrm{div}\Big(\underbrace{\frac{1}{|\nabla u|}}_{\text{diffusivity !}} \nabla u \Big).$$

Image Restoration: Deblurring


Original

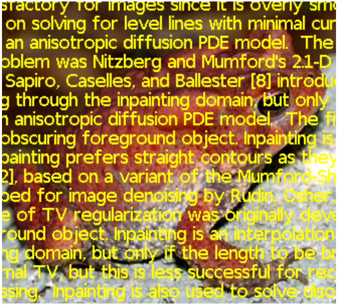

blurred and noisy


deblurred

(Goldluecke, Cremers, ICCV 2011)

Image Restoration: Inpainting

Image inpainting is a particular image restoration technique which explicitly handles (interpolate and / or extrapolate) missing data.



Corrupted

Denoised

Assume $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ a graylevel image, but only $\Omega_D \subset \Omega$ is "reliable". Then, denoising (or deblurring, etc.) should not use the f-data over $\Omega \backslash \Omega_D$. The standard TV-inpainting model is then:

$$\min_u \int_{\Omega_D} |u - f|^2 \, dx \; + \; \lambda \int_\Omega |\nabla u| \, dx.$$

Limits of TV-inpainting

TV-inpainting is a very naive interpolation technique which does not transport texture...

Overview

Solving a variational problem

There exist two strategies for solving the variational (infinite-dimension optimization) problems above, which have a form like: $\min E(u) := E_{data}(u, f) + \lambda E_{smoothness}(u)$, where E is a functional:

1. Discretize the functional, then calculate the associated (discrete) necessary optimization conditions, and compute the solution to these discrete equations (discretize-then-optimize approach).
2. Calculate the (continuous) necessary optimization conditions, discretize them and compute the solution of the discretized equations (optimize-then-discretize approach)

We next briefly recall the first approach. In the next section we will consider the second one, by introducing the continuous optimality conditions of variational problems (Euler-Lagrange equations) and how they relate to diffusion.

Discrete Denoising in 1D

A first strategy for solving variational problems is to discretize the functional, then compute the (discrete) necessary optimization conditions, and solve the latter.
Let us assume for now that f is discrete in one spatial variable, i.e. $f = \{f_1, f_2, \ldots, f_n\}$ is simply a sequence of brightness values $f_i \in \mathbb{R}$. We seek an approximation $u = \{u_1, \ldots, u_n\}$.

The data term which measures similarity of f and u can for example be written as:

$$E_{data}(u) = \frac{1}{2} \sum_{i=1}^{n} (f_i - u_i)^2,$$

which means that we measure the overall brightness difference as a sum of squared differences (SSD).

The smoothness term can for example be written as:

$$E_{smooth}(u) = \frac{1}{2} \sum_{i=1}^{n-1} (u_i - u_{i+1})^2,$$

which means that we measure the sum of squared differences for all neighboring brightness values.

Discrete Denoising in 1D

The total energy is thus:

$$E_\lambda(u) = \frac{1}{2}\sum_{i=1}^{n}(f_i - u_i)^2 + \frac{\lambda}{2}\sum_{i=1}^{n-1}(u_i - u_{i+1})^2,$$

Larger values of λ imply that the smoothness of the solution should play a bigger role.

A solution to the above denoising problem is a function $\hat{u}$ which minimizes the above energy:

$$\hat{u} = \arg\min_u E_\lambda(u).$$

Variational methods determine functions which fulfill the extremality principle:

$$\frac{dE_\lambda(u)}{du} = 0, \quad \Leftrightarrow \quad \frac{\partial E_\lambda(u)}{\partial u_i} = 0 \quad \forall i \in [1, n]$$

Discrete Denoising in 1D

The extremality condition for each pixel is therefore:

$$\frac{\partial E_\lambda(u)}{\partial u_1} = (u_1 - f_1) + \lambda(u_1 - u_2) = 0,$$

$$\frac{\partial E_\lambda(u)}{\partial u_i} = (u_i - f_i) + \lambda(2u_i - u_{i-1} - u_{i+1}) = 0, \quad \forall i \in [2, n-1],$$

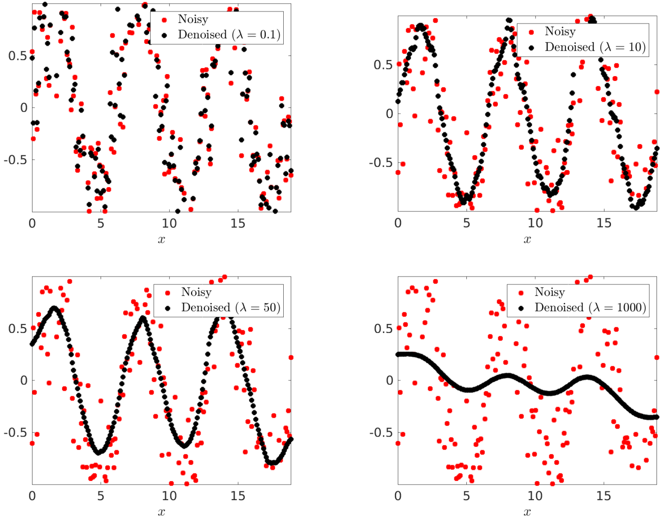
$$\frac{\partial E_\lambda(u)}{\partial u_n} = (u_n - f_n) + \lambda(u_n - u_{n-1}) = 0.$$

These conditions form a system of linear equations:

$$M_\lambda u = \begin{pmatrix} 1+\lambda & -\lambda & & & \\ -\lambda & 1+2\lambda & -\lambda & & \\ & \ddots & \ddots & \ddots & \\ & & -\lambda & 1+2\lambda & -\lambda \\ & & & -\lambda & 1+\lambda \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix}$$

Discrete Denoising in 1D

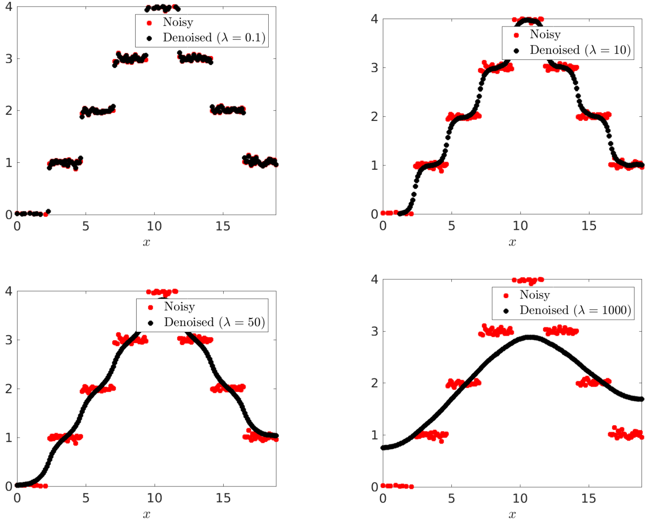
Solution of $M_\lambda u = f$ for different λ values:



$\rightarrow$ Tuning the regularization weight may be tedious

Discrete Denoising in 1D

Solution of $M_\lambda u = f$ for different λ values:



$\rightarrow$ Quadratic regularization is not always a good choice (we often prefer TV)

Discrete Denoising (d-dim.)

Index all pixels of the d-dim volume with index $i \in [1, \dots, N]$, where $N = n_1 \cdot n_2 \cdots n_d$.

Variational denoising of an image f:

$$E_\lambda(u) = \frac{1}{2}\sum_{i=1}^{N}(f_i - u_i)^2 + \frac{\lambda}{2}\sum_{i=1}^{N}\sum_{\substack{j\in\mathcal{N}(i)\\ j>i}}(u_i - u_j)^2,$$

where $\mathcal{N}(i)$ denotes the neighborhood of pixel i (typically, we consider 4-connexity in 2D).

Again E is strictly convex. The condition for (global) optimality is:

$$\frac{dE_\lambda(u)}{du_i} = (u_i - f_i) + \lambda \sum_{j\in\mathcal{N}(i)}(u_i - u_j) = 0 \qquad \forall i$$

In the higher-dimensional case, this gives rise to a large-scale linear programming problem of the form $M_\lambda u = f$.

Discrete Denoising in 2D

Solution of $M_\lambda u = f$ for different λ values:

Denoised ($\lambda = 0.1$)

Denoised ($\lambda = 10$)

Denoised ($\lambda = 50$)

Denoised ($\lambda = 1000$)

Large Scale Linear Programming

For dimensions ($d = 2, 3, ...$), the linear equation system

$$M_\lambda u = f$$

is quite large.

Most entries of the matrix M_λ are 0 (sparse matrix). Yet, its inverse is typically difficult to compute or even to store in memory.

There exist numerous standard solvers for large linear systems. The best known ones are the Jacobi method (parallelizable), Gauss-Seidel method (with its variant SOR - Successive Over-Relaxation) (faster, but non-parallelizable), and the Preconditioned Conjugate Gradient method (fastest, provided that the system matrix is positive definite).

Multi-scale / multi-grid / coarse-to-fine strategies can also be considered for acceleration.

Discrete denoising in action

Speed of convergence for different solvers:

Overview

Spatially Continuous Variational Approaches

A spatially continuous variational approach to image denoising and restoration looks as follows:

Given an image $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, find a smooth approximation $u : \Omega \to \mathbb{R}$ of this image. This can be determined by minimizing a cost function of the form:

$$E(u) = \frac{1}{2} \int \big(u(x) - f(x)\big)^2 \, dx \; + \; \frac{\lambda}{2} \int |\nabla u(x)|^2 \, dx$$

This is the spatially continuous analogue of the previously discussed discrete formulation.

Such a mapping which assigns a real number $E(u)$ to a function $u(x)$ is also referred to as a functional.

How does one minimize the above functional with respect to the function u?

Functional Minimization

A functional is a mapping E which assigns to each element of a vector-space (to each function u) an element from the underlying field (a number).

Let

$$E(u) = \int \mathcal{L}(u, u')\, dx$$

be a functional, where $u' = \frac{du}{dx}$ is the derivative of the function u. (In physics $\mathcal{L}$ is called the Lagrange density).

Example: $\mathcal{L}(u, u') = \frac{1}{2}\big(u(x) - f(x)\big)^2 + \frac{\lambda}{2}|u'(x)|^2$.

Just as with real-valued functions defined on $\mathbb{R}^n$ the necessary condition for extremality of the functional E states that the derivative with respect to u must be 0.

Yet how does one define and compute the derivative of a functional $E(u)$ with respect to the function u?

The Gâteaux Derivative

There are several ways to introduce functional derivatives. The following definition goes back to works of the French mathematician R. Gâteaux († 1914) which were published posthumously in 1919: http://archive.numdam.org/ARCHIVE/BSMF/BSMF_1919__47_/BSMF_1919__47__47_1/BSMF_1919__47__47_1.pdf

The Gâteaux derivative extends the concept of directional derivative to infinite-dimensional spaces.

The derivative of the functional $E(u)$ in direction $h(x)$ is defined as:

$$\boxed{\left.\frac{dE(u)}{du}\right|_h = \lim_{\epsilon \to 0} \frac{E(u + \epsilon h) - E(u)}{\epsilon}}$$

As in finite dimensions, this directional derivative can be interpreted as the projection of the functional gradient on the respective direction. We can therefore write:

$$\left.\frac{dE(u)}{du}\right|_h = \left\langle \frac{dE(u)}{du}, h \right\rangle = \int \frac{dE(u)}{du}(x)\, h(x)\, dx$$

The Gâteaux Derivative

For functionals of the canonical form: $E(u) = \int \mathcal{L}(u, u')\, dx$ the Gâteaux derivative is given by

$$\left.\frac{dE(u)}{du}\right|_h = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\left(E(u + \epsilon h) - E(u)\right)$$

$$= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int \left(\mathcal{L}(u + \epsilon h, u' + \epsilon h') - \mathcal{L}(u, u')\right) dx$$

$$= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int \left(\left(\mathcal{L}(u, u') + \frac{\partial \mathcal{L}}{\partial u} \epsilon h + \frac{\partial \mathcal{L}}{\partial u'} \epsilon h' + o(\epsilon^2) \right) - \mathcal{L}(u, u') \right) dx$$

$$= \int \left(\frac{\partial \mathcal{L}}{\partial u} h + \frac{\partial \mathcal{L}}{\partial u'} h' \right) dx$$

$$= \int \left(\frac{\partial \mathcal{L}}{\partial u} h - \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial u'} h \right) dx \qquad \text{(partial int., } h = 0 \text{ on boundary)}$$

$$= \int \underbrace{\left(\frac{\partial \mathcal{L}}{\partial u} - \frac{d}{dx} \frac{\partial \mathcal{L}}{\partial u'} \right)}_{\frac{dE}{du}} h(x)\, dx.$$

Euler-Lagrange Equation

Thus the derivative of the functional $E(u)$ in direction h is:

$$\frac{dE(u)}{du}\Big|_h = \int \underbrace{\left(\frac{\partial \mathcal{L}}{\partial u} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial u'}\right)}_{\frac{dE}{du}} h(x)\, dx.$$

As a necessary condition for minimality of the functional $E(u)$ the variation of E in any direction $h(x)$ must vanish. Therefore at the extremum we have:

$$\boxed{\frac{dE}{du} = \frac{\partial \mathcal{L}}{\partial u} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial u'} = 0}$$

This condition is called the Euler-Lagrange equation.

Example: For $\mathcal{L}(u, u') = \frac{1}{2}\big(u(x) - f(x)\big)^2 + \frac{\lambda}{2}|u'(x)|^2$, we get:

$$\frac{\partial \mathcal{L}}{\partial u} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial u'} = \big(u(x) - f(x)\big) - \frac{d}{dx}\big(\lambda u'(x)\big) = u - f - \lambda u'' = 0$$

Functional Minimization

The Euler-Lagrange equation is a differential equation which forms the necessary condition for minimality.

The central idea of variational methods is to compute solutions to the respective Euler-Lagrange equation.

This can be done in several ways. For example, one can discretize the function u on a set of points $\{x_1, \ldots, x_n\}$ and subsequently try to solve for the values $u(x_i)$. For quadratic cost functions, the arising set of linear equations can be solved using the discussed iterative algorithms (Jacobi, Gauss-Seidel,...). In general, however, the Euler-Lagrange equation will not be linear.

For general (non-quadratic) energies, one can start with an initial guess $u_0(x)$ of the solution and iteratively improve the solution. Such methods are called descent methods.

How can one iteratively improve a given solution?

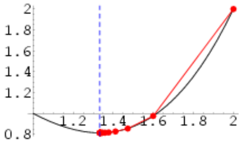
Gradient Descent

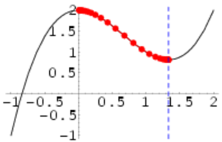
Gradient descent or steepest descent is a particular descent method where in each iteration one chooses the direction in which the energy decreases most. The direction of steepest descent is given by the negative energy gradient.

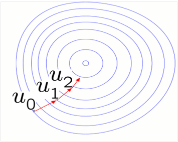
To minimize a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient descent for $f(u)$ is defined by the differential equation:

$$\begin{cases} u(0) = u_0 \\ \frac{du}{dt} = -\frac{df}{du}(u) \end{cases}$$

Discretization: $u_{t+1} = u_t - \epsilon \frac{df}{du}(u_t), \qquad t = 0, 1, 2, \ldots.$







Gradient Descent

For minimizing functionals $E(u)$, the gradient descent is done analogously.

For the functional $E(u) = \int \mathcal{L}(u, u')\, dx$, the gradient is given by:

$$\frac{dE}{du} = \frac{d\mathcal{L}}{du} - \frac{d}{dx}\frac{d\mathcal{L}}{du'}.$$

Therefore the gradient descent is given by:

$$\begin{cases} u(x,0) = u_0(x) \\ \dfrac{\partial u(x,t)}{\partial t} = -\dfrac{dE}{du} = -\dfrac{d\mathcal{L}}{du} + \dfrac{\partial}{\partial x}\dfrac{d\mathcal{L}}{du'}. \end{cases}$$

For $\mathcal{L}(u, u') = \frac{1}{2}\left(u - f\right)^2 + \frac{\lambda}{2}|u'|^2$, this means:

$$\frac{\partial u}{\partial t} = (f - u) + \lambda u'' = (f - u) + \lambda \Delta u.$$

If the gradient descent converges, i.e. $\partial_t u = -\frac{dE}{du} = 0$, then we have found a solution to the Euler-Lagrange equation.

Euler-Lagrange in 2D

Recall that in 1D we had:

$$\frac{dE}{du} = \frac{\partial \mathcal{L}}{\partial u} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial u'}.$$

Extension to 2D is as follows:

$$\boxed{\frac{dE}{du} = \frac{\partial \mathcal{L}}{\partial u} - \mathrm{div}\frac{\partial \mathcal{L}}{\nabla u}.}$$

For instance the derivative of the 2D denoising energy:

$$E(u) = \int (u-f)^2 dx + \lambda \int |\nabla u|^2 \, dx$$

is given by

$$\frac{dE}{du} = 2(u-f) - 2\,\lambda\,\mathrm{div}\nabla u = 2\,(u - f - \lambda \Delta u)$$

Diffusion with Data Term

$$E(u) = \int (f - u)^2 dx + \lambda \int |\nabla u|^2 \, dx \to \min.$$

$$E(u) = \lambda \int |\nabla u|^2 \, dx \to \min.$$

Author: D. Cremers

Diffusion as Gradient Descent

Many diffusion equations (albeit not all) can be derived as the gradient descent on a specific energy.

The energy:

$$E(u) = \int \mathcal{L}(u, \nabla u)\, dx = \frac{1}{2} \int g(x) |\nabla u(x)|^2\, dx$$

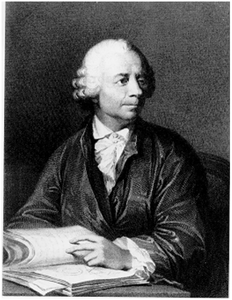
leads to the gradient descent:

$$\frac{\partial u(x,t)}{\partial t} = -\frac{dE}{du} = -\frac{\partial \mathcal{L}}{\partial u} + \text{div} \frac{\partial \mathcal{L}}{\partial \nabla u} = \text{div}\Big(g(x)\, \nabla u \Big)$$

This equation corresponds to an inhomogeneous diffusion with diffusivity $g(x)$.

In other words, the above inhomogeneous diffusion process is nothing but a steepest descent on the weighted smoothness energy.

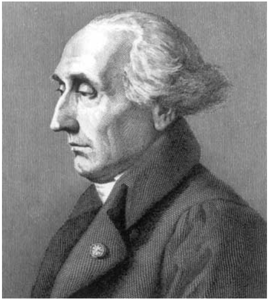
Leonhard Euler



Leonhard Euler (1707 – 1783)

- Published 886 papers and books, most of them in the last 20 years of his life. Considered the greatest mathematician of the 18th century.
- Major contributions: Euler number *e*, Euler angles, Euler formula, Euler theorem, Euler equations (for fluid flows), Euler-Lagrange equations,...
- 13 children

Joseph-Louis Lagrange



Joseph-Louis Lagrange (1736 – 1813)

- born Giuseppe Lodovico Lagrangia (in Turin). self-taught.
- With 17 years: Professor for mathematics in Turin.
- Later in Berlin (1766-1787) and Paris (1787-1813).
- 1788: *La Mécanique Analytique*.
- 1800: *Leçons sur le calcul des fonctions*.

Overview

Variational Methods and Bayesian Inference

The framework of Bayesian inference allows to systematically derive functionals for different image formation models.

Let u be the unknown true image and f the observed one, then we can write the joint probability for u and f as:

$$\mathcal{P}(u, f) = \mathcal{P}(u|f)\,\mathcal{P}(f) = \mathcal{P}(f|u)\mathcal{P}(u).$$

Rewriting this expression we obtain the Bayesian formula (Thomas Bayes 1887):

$$\mathcal{P}(u|f) = \frac{\mathcal{P}(f|u)\,\mathcal{P}(u)}{\mathcal{P}(f)}.$$

Maximum Aposteriori (MAP) estimation aims at computing the most likely solution $\hat{u}$ given f by maximizing the posterior probability $\mathcal{P}(u|f)$

$$\hat{u} = \arg\max_u \mathcal{P}(u|f) = \arg\max_u \mathcal{P}(f|u)\,\mathcal{P}(u).$$

$\mathcal{P}(f|u)$ is called the likelihood and $\mathcal{P}(u)$ the prior.

MAP Estimation in the Continuous Setting

Similarly one can define Bayesian MAP optimization in the continuous setting, where the likelihood is given by:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|^2}{2\sigma^2}dx\right),$$

and the prior is given by

$$\mathcal{P}(u) \propto \exp\left(-\lambda \int |\nabla u(x)| dx\right).$$

Thus the data term in variational methods corresponds to the likelihood, whereas the regularizer corresponds to the prior:

$$E(u) = -\log \mathcal{P}(u|f) = \int \frac{|f(x)-u(x)|^2}{2\sigma^2}dx + \lambda \int |\nabla u(x)| dx + \text{const}.$$

A systematic derivation of probability distributions on infinite-dimensional spaces requires a more formal derivation (introduction of measures etc). This is beyond our scope.

Conclusion on MAP Estimation

By invoking a Bayesian MAP rationale, one can formulate lots of computer vision problem under the form

$$\min_u D(u, f) + \lambda R(u)$$

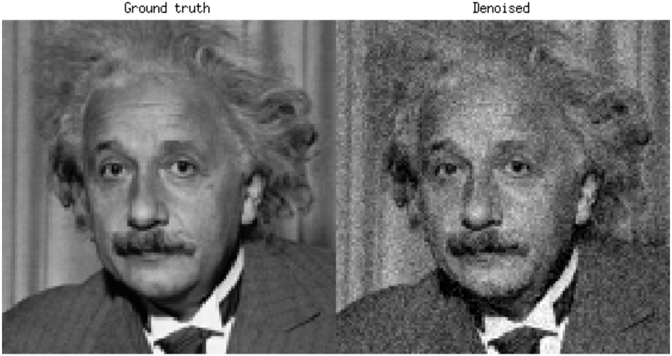
with an “automated” method for selecting D, R and λ.

A Few Classic Fidelity Terms

Gaussian noise is fine for modeling "small" perturbations:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|^2}{2\sigma^2} dx\right)$$

$$\Rightarrow \min_u \underbrace{\int |f(x)-u(x)|^2 \, \mathrm{d}x}_{:=\|f-u\|_2^2} + \lambda R(u)$$

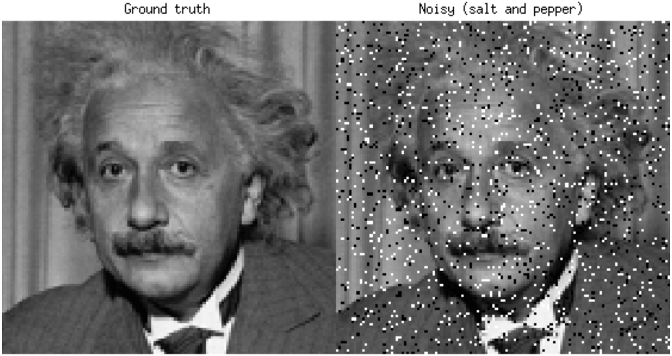


A Few Classic Fidelity Terms

Laplace noise is fine for modeling "impulsive" perturbations:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|}{\sigma}dx\right)$$

$$\Rightarrow \underbrace{\min_u \int |f(x)-u(x)|\,\mathrm{d}x}_{:=\|f-u\|_1} + \lambda R(u)$$



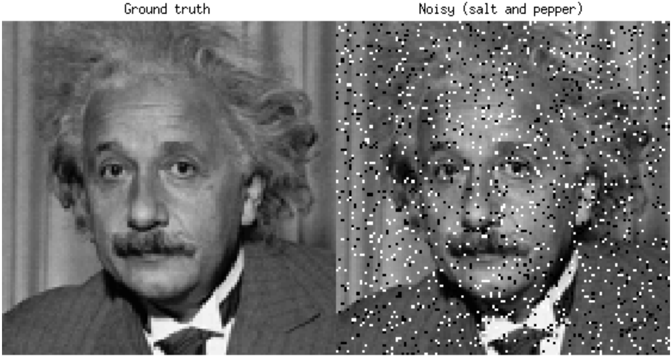
GREYC

A Few Classic Fidelity Terms

Cauchy noise is even better for modeling "impulsive" perturbations:

$$\mathcal{P}(f|u) \propto \int \frac{1}{|f(x) - u(x)|^2 + \sigma^2} dx$$

$$\Rightarrow \min_u \int \log\left(|f(x) - u(x)|^2 + \sigma^2\right) \mathrm{d}x + \lambda R(u)$$

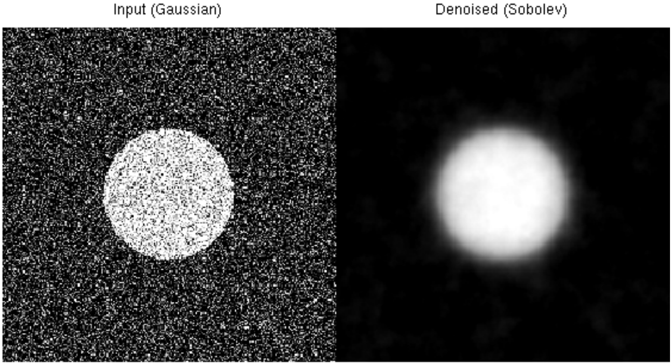


A Few Classic Regularizers

Sobolev regularization tends to favor smoothness:

$$\mathcal{P}(u) \propto \exp\left(-\lambda \int |\nabla u(x)|^2 dx\right)$$

$$\Rightarrow \min_u D(u,f) + \int |\nabla u(x)|^2 \, \mathrm{d}x$$

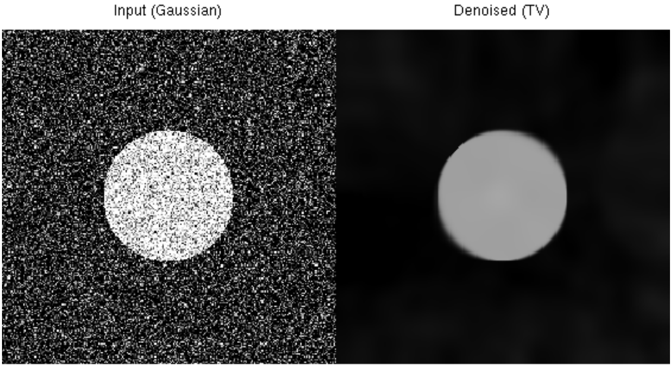


A Few Classic Regularizers

Total variation (TV) tends to favor piecewise constantness:

$$\mathcal{P}(u) \propto \exp\left(-\lambda \int |\nabla u(x)| dx\right)$$

$$\Rightarrow \min_u D(u, f) + \int \|\nabla u(x)\| \, \mathrm{d}x$$



Overview

1. Inverse Problems and Variational Methods
2. Variational Methods and Discrete Optimization
3. Variational Methods, Euler-Lagrange Equations and Diffusion
4. Variational Methods and Bayesian Inference
5. **Numerical Solving**
6. Conclusions

Gradient Descent for the L2-TV (ROF) Model

Consider the generic L2-TV (ROF) restoration problem:

$$\min_{u:\,\Omega\subset\mathbb{R}^2\to\mathbb{R}} \int_\Omega \frac{1}{2}\left((Ku)(x) - f(x)\right)^2 + \lambda|\nabla u(x)|\,\mathrm{d}x$$

(denoising: $K = \mathrm{id}$; deblurring: K = Gaussian kernel, super-resolution: K = zoom kernel, etc.)

Its first-order optimality condition is the Euler-Lagrange equation

$$K^\top(Ku - f) - \lambda\,\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right) \quad \text{over } \Omega,$$

with Neumann or Dirichlet boundary conditions on $\partial\Omega$.

Starting from $u^{t=0} = f$, optimization can be carried out by gradient descent:

$$\partial_t = -(K^\top K u) + (K^\top f) + \lambda\,\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

Main Issue in Gradient Descent for the L2-TV (ROF) Model

When implementing the gradient descent

$$\partial_t u = -(K^\top K u) + (K^\top f) + \lambda \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

one must be careful to avoid division by zero which occurs due to the factor $|\nabla u|$ (infinite diffusivity if there is no edge).

In practice, we need to smooth a bit this term:

$$\frac{1}{|\nabla u(x)|} \approx \frac{1}{|\nabla u(x)|_\epsilon} := \frac{1}{\sqrt{|\nabla u(x)|^2 + \epsilon}}$$

or

$$\frac{1}{|\nabla u(x)|} \approx \frac{1}{|\nabla u(x)|_\mu} := \frac{1}{\max\{\mu, |\nabla u(x)|\}}$$

with $\epsilon, \mu > 0$, small (e.g. 10^{-3})

The Choice of the Smoothing Matters

Example with $\epsilon = \mu = 10^{-3}$

The Choice of the Smoothing Matters

Example with $\epsilon = \mu = 10^{-2}$

The Choice of the Smoothing Matters

Example with $\epsilon = \mu = 10^{-1}$

Explicit Time Gradient Descent

We can now discretize the gradient descent equation

$$\partial_t u = -(K^\top K u) + (K^\top f) + \lambda \operatorname{div}\left(\frac{\nabla u}{|\nabla u|_\mu}\right)$$

wrt time t using forward finite differences i.e.,

$$\partial_t u = \frac{u^{(t+1)} - u^{(t)}}{\delta_t},$$

with some fixed stepsize $\delta_t > 0$.

This yields the following algorithm:

$$u^{(0)} = f$$

$$u^{(t+1)} = u^{(t)} - \delta_t \left(K^\top \left(K u^{(t)} - f \right) - \lambda \operatorname{div} \frac{\nabla u^{(t)}}{|\nabla u^{(t)}|_\mu} \right), \; t \in \{1, 2, \dots\}$$

This works, but descent has to be slow (low δ_t)

Lagged Diffusivity (Implicit Time Gradient Descent)

To make things more stable, we usually prefer to freeze only the diffusivity during descent, i.e.:

$$u^{(0)} = f$$

$$u^{(t+1)} = u^{(t)} - \delta_t \left(K^\top \left(K u^{(t+1)} - f \right) - \lambda \operatorname{div} \frac{\nabla u^{(t+1)}}{|\nabla u^{(t)}|_\mu} \right), \; t \in \{1, 2, \ldots\},$$

which requires a linear system to be solved at each update:

$$\left(\mathrm{id} + \delta_t K^\top K - \delta_t \lambda \operatorname{div} \left(\frac{1}{|\nabla u^{(t)}|_\mu} \nabla \right) \right) u^{(t+1)} = u^{(t)} + \delta_t K^\top f$$

Typically much larger stepsizes are allowed, which makes things way faster and removes the need for tedious tuning (or linesearch).

The Choice of the Stepsize Matters

Example with $\delta_t = 0.02$

The Choice of the Stepsize Matters

Example with $\delta_t = 0.2$

The Choice of the Stepsize Matters

Example with $\delta_t = 2$

Gradient Descent Process for the Inpainting + Denoising Task

Overview

1. Inverse Problems and Variational Methods
2. Variational Methods and Discrete Optimization
3. Variational Methods, Euler-Lagrange Equations and Diffusion
4. Variational Methods and Bayesian Inference
5. Numerical Solving
6. **Conclusions**

Recap

What we have seen so far:

- Lots of low-level imaging problems can be formulated as variational problems (Ch. 1)
- Image filtering $\equiv$ Diffusion (Ch. 2)
- Diffusion $\equiv$ Gradient descent on the variational problem (Ch. 3)
- Variational modeling $\equiv$ Bayesian rationale (Ch. 3)

$\Rightarrow$ Unifying PDEs, optimization, functional analysis and statistics for low-level imaging

Next: Higher-level applications (Poisson editing, 3D-reconstruction, image segmentation...) (Ch. 4)