Foundations of Variational Image Analysis
Lecture Notes, HCI, 4.10.2011

Chapter 1

Variational Analysis

Euler-Lagrange Equations and Linear Inverse Problems

Bastian Goldlücke
Computer Vision Group
Technical University of Munich

Overview

Overview

A simple (but important) example: Denoising

The TV-$\mathcal{L}^2$ (ROF) model, Rudin-Osher-Fatemi 1992

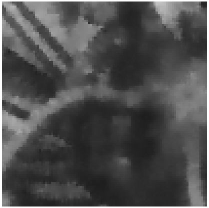
For a given noisy input image f, compute

$$\operatorname*{argmin}_{u \in \mathcal{L}^2(\Omega)} \Bigg[\underbrace{\int_\Omega |\nabla u|_2 \, dx}_{\text{regularizer / prior}} + \underbrace{\frac{1}{2\lambda} \int_\Omega (u - f)^2 \, dx}_{\text{data / model term}} \Bigg].$$

Note: In Bayesian statistics, this can be interpreted as a MAP estimate for Gaussian noise.

Original

Noisy



Reconstruction

Reminder: the space $\mathcal{L}^2(\Omega)$

Definition

Let $\Omega \subset \mathbb{R}^n$ open. The space $\mathcal{L}^2(\Omega)$ of square-integrable functions is defined as

$$\mathcal{L}^2(\Omega) := \left\{ u : \Omega \to \mathbb{R} \; : \; \left(\int_\Omega |u|^2 \, dx \right)^{\frac{1}{2}} < \infty \right\}.$$

Reminder: the space $\mathcal{L}^2(\Omega)$

- The functional
$$\|u\|_2 := \left(\int_\Omega |u|^2 \, dx \right)^{\frac{1}{2}}$$
is a norm on $\mathcal{L}^2(\Omega)$, with which it becomes a Banach space.
- The norm arises from the inner product
$$(u, v) \mapsto \int_\Omega uv \, dx$$
if you set $\|u\|_2 := \sqrt{(u, u)}$. Thus, $\mathcal{L}^2(\Omega)$ is in fact a Hilbert space. It is one of the most simple examples for an infinite dimensional Hilbert space.
- In the following, we assume functions to be in $\mathcal{L}^2(\Omega)$, and convergence, continuity etc. is defined with respect to the above norm.

$\mathbb{R}^n$ vs. $\mathcal{L}^2(\Omega)$

	$\mathcal{V} = \mathbb{R}^n$	$\mathcal{V} = \mathcal{L}^2(\Omega)$
Elements	finitely many components $x_i, 1 \leq i \leq n$	infinitely many "components" $u(x), x \in \Omega$
Inner Product	$(x, y) = \sum_{i=1}^n x_i y_i$	$(u, v) = \int_\Omega uv \, dx$
Norm	$\lvert x \rvert_2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\lVert u \rVert_2 = \left(\int_\Omega \lvert u \rvert^2 \, dx \right)^{\frac{1}{2}}$

Derivatives of a functional $E : \mathcal{V} \to \mathbb{R}$

Gradient (Fréchet)	$dE(x) = \nabla E(x)$	$dE(u) = ?$
Directional (Gâteaux)	$\delta E(x; h) = \nabla E(x) \cdot h$	$\delta E(u; h) = ?$
Condition for minimum	$\nabla E(\hat{x}) = 0$	$?$

Gâteaux differential

Definition

Let $\mathcal{V}$ be a vector space, $E : \mathcal{V} \to \mathbb{R}$ a functional, $u, h \in \mathcal{V}$. If the limit

$$\delta E(u; h) := \lim_{\alpha \to 0} \frac{1}{\alpha} \left(E(u + \alpha h) - E(u) \right)$$

exists, it is called the Gâteaux differential of E at u with increment h.

- The Gâteaux differential can be though of as the directional derivative of E at u in direction h.
- A classical term for the Gâteaux differential is “variation of E”, hence the term “variational methods”. You test how the functional “varies” when you go into direction h.

The variational principle

The variational principle is a generalization of the necessary condition for extrema of functions on $\mathbb{R}^n$.

Theorem (variational principle)

If $\hat{u} \in \mathcal{V}$ is an extremum of a functional $E : \mathcal{V} \to \mathbb{R}$, then

$$\delta E(\hat{u}; h) = 0 \text{ for all } h \in \mathcal{V}.$$

For a proof, note that if $\hat{u}$ is an extremum of E, then 0 must be an extremum of the real function

$$t \mapsto E(\hat{u} + th)$$

for all h.

Euler-Lagrange equation

The Euler-Lagrange equation is a PDE which has to be satisfied by an extremal point $\hat{u}$. A ready-to-use formula can be derived for energy functionals of a specific, but very common form.

Theorem

Let $\hat{u}$ be an extremum of the functional $E : \mathcal{C}^1(\Omega) \to \mathbb{R}$, and E be of the form

$$E(u) = \int_\Omega L(u, \nabla u, x)\,\mathrm{d}x,$$

with $L : \mathbb{R} \times \mathbb{R}^n \times \Omega \to \mathbb{R}, (a, b, x) \mapsto L(a, b, x)$ continuously differentiable. Then $\hat{u}$ satisfies the Euler-Lagrange equation

$$\partial_a L(u, \nabla u, x) - \mathrm{div}_x \left[\nabla_b L(u, \nabla u, x)\right] = 0,$$

where the divergence is computed with respect to the location variable x, and

$$\partial_a L := \frac{\partial L}{\partial a}, \nabla_b L := \left[\frac{\partial L}{\partial b_1} \dots \frac{\partial L}{\partial b_n}\right]^T.$$

Fundamental lemma of variational calculus

The derivation of the Euler-Lagrange equation requires two theorems:

- The DuBois-Reymond lemma, the most general form of the "fundamental lemma of variational calculus",
- The divergence theorem of Gauss, which can be thought of as a form of "integration by parts" for higher-dimensional spaces.

DuBois-Reymond lemma

Take $u \in \mathcal{L}^1_{\text{loc}}(\Omega)$. If

$$\int_\Omega u(x)h(x)\,\mathrm{d}x = 0$$

for every test function $h \in \mathcal{C}_c^\infty(\Omega)$, then $u = 0$ almost everywhere.

Derivation of Euler-Lagrange equation (1)

Let $h \in \mathcal{C}_c^\infty(\Omega)$ be a test function. The central idea for deriving the Euler-Lagrange equation is to compute the Gâteaux derivative of E at u in direction h, and write it in the form

$$\delta E(u; h) = \int_\Omega \phi_u h \, \mathrm{d}x,$$

with a function $\phi_u : \Omega \to \mathbb{R}$. Since at an extremum, this expression is zero for arbitrary test functions h, the Euler-Lagrange equation $\phi_u = 0$ will then follow from the fundamental lemma.

Note: The equality above shows that the function ϕ_u is the generalization of the gradient, since directional derivatives are computed via the linear map

$$h \mapsto (\phi_u, h).$$

The function ϕ_u represents the so-called Fréchet derivative of E at u.

Divergence theorem of Gauss

Divergence theorem (Gauss)

Suppose $\Omega \subset \mathbb{R}^n$ is compact with piecewise smooth boundary, $\mathbf{n} : \partial\Omega \to \mathbb{R}^n$ the outer normal of Ω and $\mathbf{p} : \mathbb{R}^n \to \mathbb{R}^n$ a continuously differentiable vector field, defined at least in a neighbourhood of Ω. Then

$$\int_\Omega \operatorname{div}(\mathbf{p})\, dx = \oint_{\partial\Omega} \mathbf{p} \cdot \mathbf{n}\, ds.$$

Corollary: integration by parts

If in addition, $u : \Omega \to \mathbb{R}$ is a differentiable scalar function, then

$$\int_\Omega \nabla u \cdot \mathbf{p}\, dx = -\int_\Omega u \cdot \operatorname{div}(\mathbf{p})\, dx + \oint_{\partial\Omega} u\mathbf{p} \cdot \mathbf{n}\, ds.$$

Derivation of Euler-Lagrange equation (2)

The Gâteaux derivative of E at u in direction h is

$$\delta E(u;h) = \lim_{\alpha \to 0} \frac{1}{\alpha} \int_\Omega L(u+\alpha h, \nabla(u+\alpha h), x) - L(u, \nabla u, x)\, dx.$$

Because of the assumptions on L, we can take the limit below the integral and apply the chain rule to get

$$\delta E(u;h) = \int_\Omega \partial_a L(u, \nabla u, x) h + \nabla_b L(u, \nabla u, x) \cdot \nabla h \, dx.$$

Applying integration by parts to the second part of the integral with $\mathbf{p} = \nabla_b L(u, \nabla u, x)$, noting $h\big|_{\partial\Omega} = 0$, we get

$$\delta E(u;h) = \int_\Omega \Big(\partial_a L(u, \nabla u, x) - \mathrm{div}_x \left[\nabla_b L(u, \nabla u, x) \right] \Big) \cdot h \, dx.$$

This is the desired expression, from which we can directly see the definition of ϕ_u.

Overview

The ROF functional

The Rudin-Osher-Fatemi (ROF) model

Given an image $f \in \mathcal{L}^2(\Omega)$ and a smoothing parameter $\lambda > 0$, compute a denoised image

$$\hat{u} \in \operatorname*{argmin}_{u \in \mathcal{L}^2(\Omega)} \int_\Omega |\nabla u|_2 + \frac{1}{2\lambda}(u - f)^2 \, dx$$

- The model was introduced in the (now famous) paper “Nonlinear total variation based noise removal algorithms” by Leonid Rudin, Stanley Osher and Emad Fatemi in 1992, and interestingly appeared in *Physica D*, a specialized journal for “nonlinear phenomena” in natural sciences.
- Note that in the notation above, u is required to be differentiable. This will be remedied later.

Euler-Lagrange equation for the ROF functional I

- The ROF functional is of the form

$$E(u) = \int_{\Omega} L(u, \nabla u, x)\,\mathrm{d}x$$

with

$$L(a, b, x) := \sqrt{b_1^2 + b_2^2} + \frac{1}{2\lambda}(a - f(x)).$$

- The problem is that the norm is not differentiable at $b = 0$. Thus, one can only compute gradient descent for an approximated L_ϵ for a regularization parameter $\epsilon > 0$:

$$L_\epsilon(a, b, x) := \underbrace{\sqrt{b_1^2 + b_2^2 + \epsilon}}_{=:|b|_\epsilon} + \frac{1}{2\lambda}(a - f(x))^2.$$

Euler-Lagrange equation for the ROF functional II

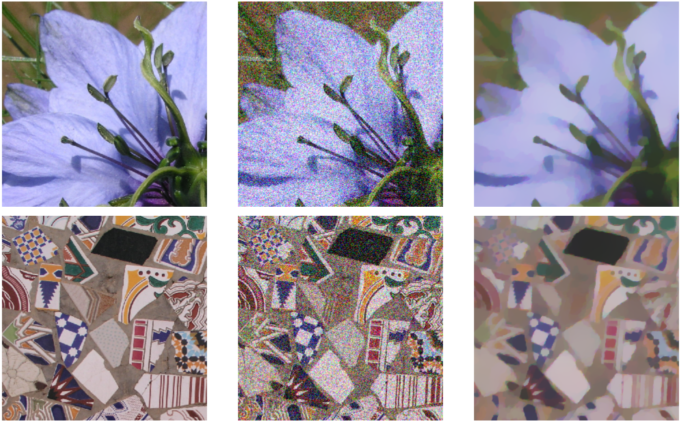
- The approximation L_ϵ is differentiable everywhere, with

$$\partial_a L_\epsilon(u, \nabla u, x) = \frac{1}{\lambda}(u(x) - f(x))$$
$$\nabla_b L_\epsilon(u, \nabla u, x) = \frac{\nabla u(x)}{|\nabla u(x)|_\epsilon}$$

- Thus, according to the theorem, the Euler-Lagrange equation of the ROF functional is given by

$$-\text{div}\left(\frac{\nabla u}{|\nabla u|_\epsilon}\right) + \frac{1}{\lambda}(u - f) = 0.$$

More ROF examples



Original *Noisy* *Solution*

Inpainting problem

In the inpainting problem, we try to recover missing areas of a damaged picture as plausibly as possible from the known areas around them.

Damaged image f

Recovered image u

Technically, we are given a damaged region $\Gamma \subset \Omega$, and a partial image $f : \Omega \setminus \Gamma \to \mathbb{R}$ defined only outside the damaged region. We want to recover $u : \Omega \to \mathbb{R}$ with $u|_{\Omega \setminus \Gamma} = f$.

Object removal

Once we have an inpainting algorithm, we can also employ it to remove unwanted regions in an image.

Original image u

Removed region Γ

Inpainted result

TV inpainting

The idea in TV inpainting is that the missing regions are filled in by minimizing the total variation, while keeping close to the original image in the known regions.

TV inpainting model

$$\underset{u \in \mathcal{L}^2(\Omega)}{\operatorname{argmin}} \left[\int_\Omega \lambda \, |\nabla u|_2 + (1 - 1_\Gamma)(u - f)^2 \, dx \right],$$

where 1_Γ is the characteristic function of Γ, i.e.

$$1_\Gamma(x) = \begin{cases} 1 & \text{if } x \in \Gamma, \\ 0 & \text{otherwise.} \end{cases}$$

The constant $\lambda > 0$ is chosen small so that smoothing is minimal outside of the inpainting region.

It looks the same as the ROF model, but there is a factor before the data term which depends on the location in the image.

TV inpainting results ("cartoon" images)

Original *Damaged* *Inpainted result*

The results are ok given the simplicity of the model, but nothing to be really proud of.

TV inpainting results (“textured” images)

Original *Damaged* *Inpainted*

- TV inpainting is unconvincing for highly textured images if the missing regions are larger. The reason is that no structure is inferred from surrounding regions, and only boundary values of Γ are taken into account.
- If you are looking for a variational model for inpainting, look out for papers on non-local TV by Osher et al.

Convolution

Definition

Let $b, u : \mathbb{R}^n \to \mathbb{R}$. The convolution $b * u$ of b and u is also a function $\mathbb{R}^n \to \mathbb{R}$. It is defined point-wise as

$$(b * u)(x) := \int_{\mathbb{R}^n} b(y)u(x - y)\,\mathrm{d}y.$$

Remarks:

- We define convolutions with functions defined only on $\Omega \subset \mathbb{R}^n$ by first extending the function to the full space via setting it to zero outside of Ω.
- If $b \in \mathcal{L}^1(\mathbb{R}^n)$ and $u \in \mathcal{L}^p(\mathbb{R}^n)$, then the convolution $b * u$ will also be in $\mathcal{L}^p(\mathbb{R}^n)$.

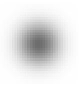
Algebraic properties of convolution

- Commutativity:
$$b * u = u * b$$
- Associativity:
$$b * (u * v) = (b * u) * v$$
- Distributivity:
$$b * (u + v) = b * u + b * v$$
- Associativity with scalar multiplication:
$$\lambda(b * u) = \lambda b * u = b * (\lambda u)$$

Blurring

The convolution $b * u$ with a kernel b of total mass 1 can be interpreted as a blurring operation.

Example: Gaussian blur (isotropic)



*

=

Example: Motion blur for diagonal motion (anisotropic)

*

=

TV deblurring model

The idea is that you observe an image f, which results from u to be blurred and contaminated with Gaussian noise. Thus, a useful model to recover u is to use an $\mathcal{L}^2$ distance in the data term. As a regularizer, we choose TV again.

TV deblurring

Given an image f which is noisy and blurred with blur kernel b. In order to recover the original u, we solve

$$\operatorname*{argmin}_{u \in \mathcal{L}^2(\Omega)} \left[\int_\Omega |\nabla u|_2 + \frac{1}{2\lambda} \int_\Omega (b * u - f)^2 \, \mathrm{d}x \right].$$

Euler-Lagrange equations

We already know how to compute derivatives for the common regularizers. Thus, we only need the derivative for the new data term.

Proposition

Let $E(u) := \int_\Omega (b * u - f)^2$. Then the Gâteaux derivative of E is given by

$$\delta E(u; h) = \int_\Omega \left[2\bar{b} * (b * u - f) \right] h \, dx,$$

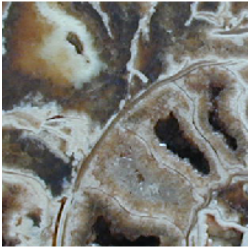
where $\bar{b}$ is the kernel adjoint to b defined by $\bar{b}(x) = b(-x)$.

For the proof, just start with computing the Gâteaux derivative as usual. At some point, you will need to "shift" a convolution away from h, similar as we shifted a derivative for (3.16) with Gauss theorem. For this, you need to make use of the fact that convolution with $\bar{b}$ is "adjoint" to convolution with b, which means that

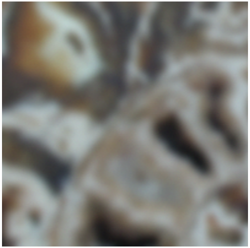
$$\int_\Omega (\bar{b} * g) h \, dx = \int_\Omega g (b * h) \, dx.$$

Try to proof this as an exercise.

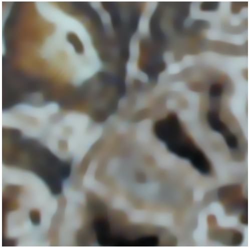
Example (1)



Original



Blurred



Solution

Of course, fine details removed by the blurring process are forever lost and cannot be recovered. However, we can reconstruct a sharper image and the location of image edges.

Example (2)

Original *Blurred* *Solution*

Generalization: Linear Inverse Problems

Proposition

Let $E(u) := \int_\Omega (Au - f)^2$. Then the Gâteaux derivative of E is given by

$$\delta E(u; h) = \int_\Omega [2A^*(Au - f)]\, h \, dx,$$

where A^* is the adjoint operator of A, i.e.

$$\langle u, A^* v \rangle = \langle Au, v \rangle \text{ for all } u, v \in \mathcal{L}^2(\Omega).$$

- For the proof, just start with the definition of the Gâteaux derivative as usual. Use the defining equation for A^* to “shift” the operator A away from h.
- Note that this shows that the adjoint of a convolution operation is the convolution with the adjoint kernel.

Summary

- Variational calculus deals with functionals on infinite-dimensional vector spaces.
- Minima are characterized by the variational principle, which leads to the Euler-Lagrange equation for a large class of functionals.
- The left-hand side of the Euler-Lagrange equation yields the Fréchet derivative of the functional.
- We have discussed the classical examples: denoising, inpatinting, deblurring and general linear inverse problems.

Open questions

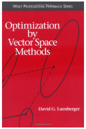
- How can we compute solution to the Euler-Lagrange equation?
- The regularizer of the ROF functional is

$$\int_{\Omega} |\nabla u|_2 \, dx,$$

  which requires u to be differentiable. Yet, we are looking for minimizers in $\mathcal{L}^2(\Omega)$. It is necessary to generalize the definition of the regularizer.
- The total variation is not a differentiable functional, so the variational principle is not applicable. We need a theory for convex, but not differentiable functionals.

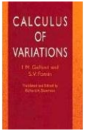
References

Variational methods



Luenberger,
“Optimization by Vector Space Methods”,
Wiley 1969.

- Elementary introduction of optimization on Hilbert and Banach spaces.
- Easy to read, many examples from other disciplines, in particular economics.



Gelfand and Fomin,
“Calculus of Variations”,
translated 1963 (original in Russian).

- Classical introduction of variational calculus, somewhat outdated terminology, inexpensive and easy to get
- Historically very interesting, lots of non-computer-vision applications (classical geometric problems, Physics: optics, mechanics, quantum mechanics, field theory)