Variational Methods in Computer Vision
ICCV Tutorial, 6.11.2011

Chapter 1

Introduction

Mathematical Foundations

Daniel Cremers and Bastian Goldlücke
Computer Vision Group
Technical University of Munich

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology

Overview

1. **Variational Methods**
   - Introduction
   - Convex vs. non-convex functionals
   - Archetypical model: ROF denoising
   - The variational principle
   - The Euler-Lagrange equation
2. **Total Variation and Co-Area**
   - The space $\mathcal{BV}(\Omega)$
   - Geometric properties
   - Co-area
3. **Convex analysis**
   - Convex functionals
   - Constrained Problems
   - Conjugate functionals
   - Subdifferential calculus
   - Proximation and implicit subgradient descent
4. **Summary**

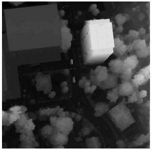
Overview

Fundamental problems in computer vision

Image labeling problems

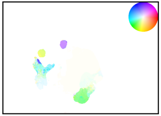

Segmentation and Classification





Stereo

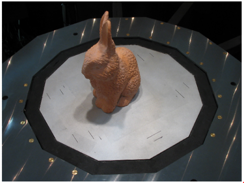


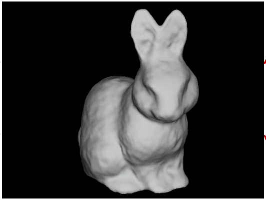
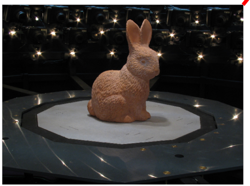
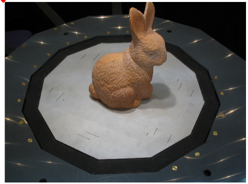


Optic flow

Fundamental problems in computer vision

3D Reconstruction



Variational methods

Unifying concept: variational approach

Variational methods

Unifying concept: variational approach

Problem solution is the minimizer of an energy functional E,

$$\operatorname*{argmin}_{u \in \mathcal{V}} E(u).$$

Variational methods

Unifying concept: variational approach

Problem solution is the minimizer of an energy functional E,

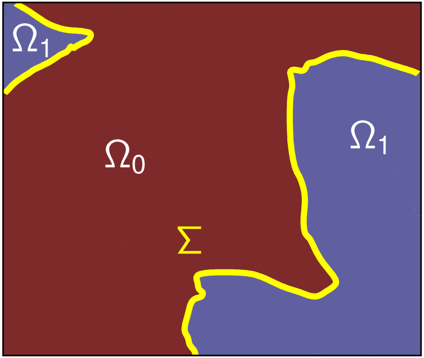
$$\operatorname*{argmin}_{u \in \mathcal{V}} E(u).$$

In the variational framework, we adopt a
continuous world view.

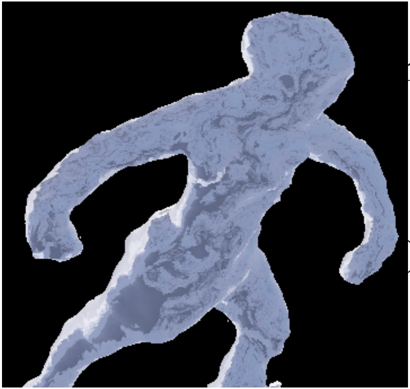
Images are functions

A greyscale image is a real-valued function
$u : \Omega \to \mathbb{R}$ on an open set $\Omega \subset \mathbb{R}^2$.

Surfaces are manifolds



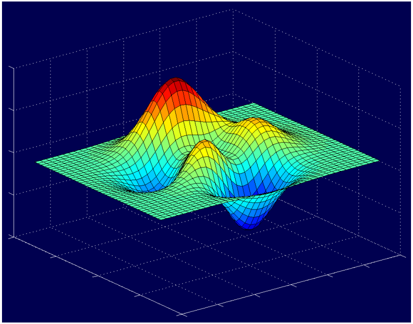
2D



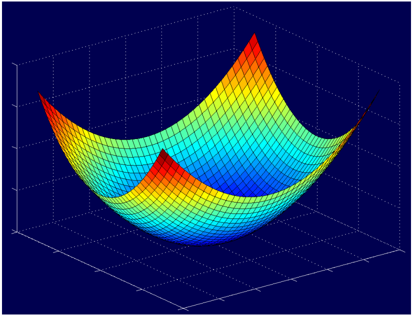
3D

Volume usually modeled as the level set $\{x \in \Omega : u(x) = 1\}$ of a binary function $u : \Omega \to \{0, 1\}$.

Convex versus non-convex energies

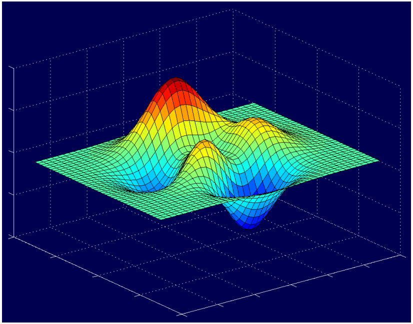


non-convex energy

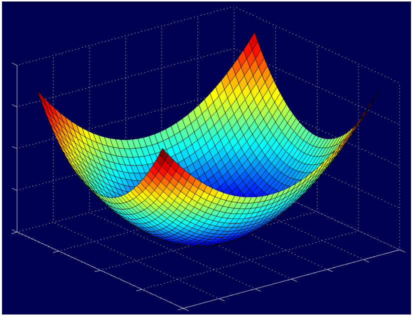


convex energy

Convex versus non-convex energies



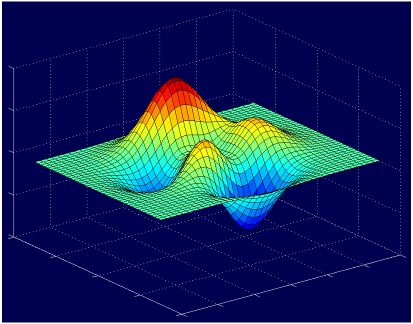
non-convex energy



convex energy

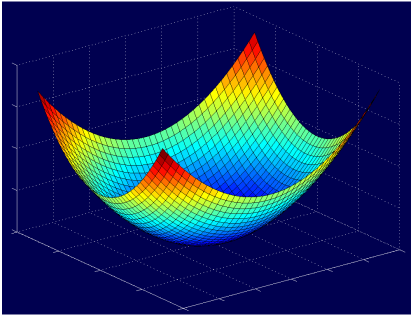
- Cannot be globally minimized
- Efficient global minimization

Convex versus non-convex energies



non-convex energy

- Cannot be globally minimized
- Realistic modeling

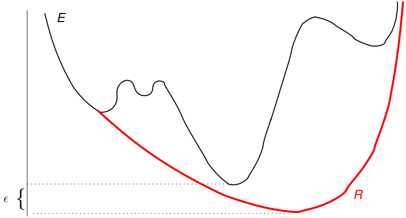


convex energy

- Efficient global minimization
- Often unrealistic models

Convex relaxation

Convex relaxation: best of both worlds?



- Start with realistic *non-convex* model energy E
- *Relax* to convex lower bound R, which can be efficiently minimized
- Find a (hopefully small) *optimality bound* ϵ to estimate quality of solution.

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

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

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

Original

Noisy

Result, $\lambda = 2$

$\mathbb{R}^n$ 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	$(x, y) = \sum_{i=1}^n x_i y_i$	$(u, v) = \int_\Omega uv \, \mathrm{d}x$
Norm	$\lvert x \rvert_2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\lVert u \rVert_2 = \left(\int_\Omega \lvert u \rvert^2 \, \mathrm{d}x \right)^{\frac{1}{2}}$

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.

Derivation of Euler-Lagrange equation (1)

Method:

- Compute the Gâteaux derivative of E at u in direction h, and write it in the form

$$\delta E(u; h) = \int_\Omega \phi_u h \, dx,$$

with a function $\phi_u : \Omega \to \mathbb{R}$ and a test function $h \in \mathcal{C}_c^\infty(\Omega)$.

- At an extremum, this expression must be zero for arbitrary test functions h, thus (due to the "duBois-Reymond Lemma") you get the condition

$$\phi_u = 0.$$

This is the Euler-Lagrange equation of the functional E.

- Note: the form above is in analogy to the finite-dimensional case, where the gradient satisfies $\delta E(x; h) = \langle \nabla E(x), \cdot h \rangle$.

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

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

Derivation of Euler-Lagrange equation (2)

The Gâteaux derivative of E at u in direction h is

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

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

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

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

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

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

Open questions

- The regularizer of the ROF functional is

$$\int_\Omega |\nabla u|_2 \, \mathrm{d}x,$$

  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, which will lead to the total variation in the next section.
- 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.

Overview

Definition of the total variation

- Let $u \in \mathcal{L}^1_{\text{loc}}(\Omega)$. Then the total variation of u is defined as

$$J(u) := \sup \left\{ -\int_\Omega u \cdot \operatorname{div}(\xi)\, dx \;:\; \xi \in \mathcal{C}^1_c(\Omega, \mathbb{R}^n), \|\xi\|_\infty \leq 1 \right\}.$$

- The space $\mathcal{BV}(\Omega)$ of functions of bounded variation is defined as

$$\mathcal{BV}(\Omega) := \left\{ u \in \mathcal{L}^1_{\text{loc}}(\Omega) \;:\; J(u) < \infty \right\}.$$

An idea why this is the same as before: the norm of the gradient $|\nabla u(x)|_2$ below the integral for a fixed point x can be written as

$$|\nabla u(x)|_2 = \sup_{\xi \in \mathbb{R}^n : |\xi|_2 \leq 1} \nabla u(x) \cdot \xi.$$

We can then use Gauss' theorem again to shift the gradient from u to a divergence on ξ.

Convexity and lower-semicontinuity

Below are the main analytical properties of the total variation. It also enjoys a number of interesting geometrical relationships, which will be explored next.

Proposition

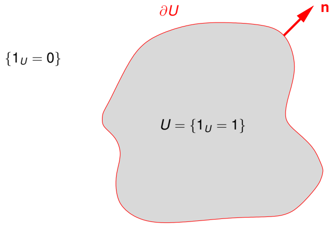
- J is a semi-norm on $\mathcal{BV}(\Omega)$, and it is convex on $\mathcal{L}^2(\Omega)$.
- J is lower semi-continuous on $\mathcal{L}^2(\Omega)$, i.e.

$$\|u_n - u\|_2 \to 0 \implies J(u) \leq \liminf_{u_n} J(u_n).$$

The above can be shown immediately from the definition, lower semi-continuity requires Fatou's Lemma.

Lower semi-continuity is important for the existence of minimizers, see next section.

Characteristic functions of sets



Let $U \subset \Omega$. Then the characteristic function of U is defined as

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

Notation

If $u : \Omega \to \mathbb{R}$ then $\{f = 0\}$ is a short notation for the set

$$\{x \in \Omega \ : \ f(x) = 0\} \subset \Omega.$$

Similar notation is used for inequalities and other properties.

Total variation of a characteristic function

We now compute the TV of the characteristic function of a "sufficiently nice" set $U \subset \Omega$, with a $\mathcal{C}^1$-boundary.

Remember: to compute the total variation, one maximizes over all vector fields $\xi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^n), \|\xi\|_\infty \leq 1$:

$$
\begin{aligned}
-\int_\Omega 1_U \cdot \operatorname{div}(\xi)\,dx &= -\int_U \operatorname{div}(\xi)\,dx \\
&= \int_{\partial U} \mathbf{n} \cdot \xi \, ds \text{ (Gauss' theorem)}
\end{aligned}
$$

The expression is maximized for any vector field with $\xi|_{\partial U} = \mathbf{n}$, hence

$$
J(1_U) = \int_{\partial U} ds = \mathcal{H}^{n-1}(\partial U).
$$

Here, $\mathcal{H}^{n-1}$, is the $(n-1)$-dimensional Haussdorff measure, i.e. the length in the case $n=2$, or area for $n=3$.

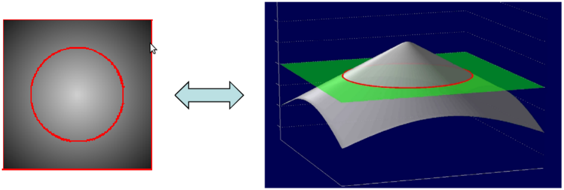
The co-area formula

The co-area formula in its geometric form says that the total variation of a function equals the integral over the $(n-1)$-dimensional area of the boundaries of all its lower level sets. More precisely,

Theorem (co-area formula)

Let $u \in \mathcal{BV}(\Omega)$. Then

$$J(u) = \int_{-\infty}^{\infty} J(1_{\{u \leq t\}})\,\mathrm{d}t$$



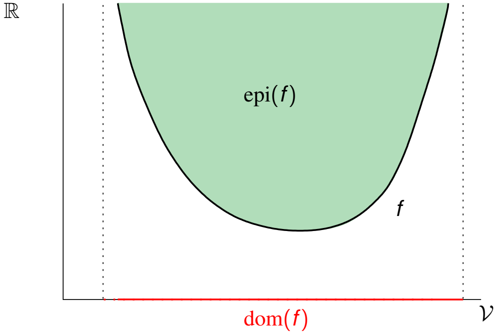
Overview

The epigraph of a functional

Definition

The epigraph $\operatorname{epi}(f)$ of a functional $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ is the set "above the graph", i.e.

$$\operatorname{epi}(f) := \{(x, \mu) \ : \ x \in \mathcal{V} \text{ and } \mu \geq f(x)\}.$$



Convex functionals

We choose the geometric definition of a convex function here because it is more intuitive, the usual algebraic property is a simple consequence.

Definition

- A functional $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ is called proper if $f \neq \infty$, or equivalently, the epigraph is non-empty.
- A functional $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ is called convex if epi(f) is a convex set.
- The set of all proper and convex functionals on $\mathcal{V}$ is denoted conv($\mathcal{V}$).

The only non-proper function is the constant function $f = \infty$. We exclude it right away, otherwise some theorems become cumbersome to formulate. From now on, every functional we write down will be proper.

Extrema of convex functionals

Convex functionals have some very important properties with respect to optimization.

Proposition

Let $f \in \text{conv}(\mathcal{V})$. Then

- the set of minimizers $\text{argmin}_{x \in \mathcal{V}} f(x)$ is convex (possibly empty).
- if $\hat{x}$ is a local minimum of f, then $\hat{x}$ is in fact a global minimum, i.e. $\hat{x} \in \text{argmin}_{x \in \mathcal{V}} f(x)$.

Both can be easily deduced from convexity of the epigraph.

Lower semi-continuity and closed functionals

Lower semi-continuity is an important property for convex functionals, since together with coercivity it guarantees the existence of a minimizer. It has an intuitive geometric interpretation.

Definition

Let $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ be a functional. Then f is called closed if $\mathrm{epi}(f)$ is a closed set.

Proposition (closedness and lower semi-continuity)

For a functional $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$, the following two are equivalent:

- f is closed.
- f is lower semi-continuous, i.e.

$$f(x) \leq \liminf_{x_n \to x} f(x_n)$$

for any sequence (x_n) which converges to x.

An existence theorem for a minimum

Definition

Let $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ be a functional. Then f is called coercive if it is "unbounded at infinity". Precisely, for any sequence $(x_n) \subset \mathcal{V}$ with $\lim \|x_n\| = \infty$, we have $\lim f(x_n) = \infty$.

Theorem

Let f be a closed, coercive and convex functional on a Banach space $\mathcal{V}$. Then f attains a minimum on $\mathcal{V}$.

The requirement of coercivity can be weakened, a precise condition and proof is possible to formulate with the subdifferential calculus. On Hilbert spaces (and more generally, the so-called "reflexive" Banach spaces), the requirements of "closed and convex" can be replaced by "weakly lower semi-continuous". See [Rockafellar] for details.

Examples

- The function $x \mapsto \exp(x)$ is convex, lower semi-continuous but not coercive on $\mathbb{R}$. The infimum 0 is not attained.
- The function
$$x \mapsto \begin{cases} \infty & \text{if } x \leq 0 \\ x^2 & \text{if } x > 0 \end{cases}$$
is convex, coercive, but not closed on $\mathbb{R}$. The infimum 0 is not attained.
- The functional of the ROF model is closed and convex. It is also coercive on $\mathcal{L}^2(\Omega)$: from the inverse triangle inequality,
$$|\|u\|_2 - \|f\|_2| \leq \|u - f\|_2 \,.$$
Thus, if $\|u_n\|_2 \to \infty$, then
$$E(u_n) \geq \|u_n - f\|_2 \geq |\|u_n\|_2 - \|f\|_2| \to \infty.$$
Therefore, there exists a minimizer of ROF for each input $f \in \mathcal{L}^2(\Omega)$.

The Indicator Function of a Set

Definition

For any subset $S \subset \mathcal{V}$ of a vector space, the indicator function $\delta_S : \mathcal{V} \to \mathbb{R} \cup \infty$ is defined as

$$\delta_S(x) := \begin{cases} \infty & \text{if } x \neq S, \\ 0 & \text{if } x \in S. \end{cases}$$

Indicator functions give examples for particularly simple convex functions, as they have only two different function values.

Proposition (convexity of indicator functions)

S is a convex set if and only if δ_S is a convex function.

The proposition is easy to prove (exercise). Note that by convention, $r < \infty$ for all $r \in \mathbb{R}$.

Constrained Problems

Suppose you want to find the minimizer of a convex functional $f : C \to \mathbb{R}$ defined on a convex set $C \subset \mathcal{V}$. You can always exchange that with an unconstrained problem which has the same minimizer: introduce an extended function

$$\tilde{f} : \mathcal{V} \to \mathbb{R}, \qquad \tilde{f}(x) := \begin{cases} f(x) & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases}$$

Then

$$\operatorname*{argmin}_{x \in C} f(x) = \operatorname*{argmin}_{x \in \mathcal{V}} \tilde{f}(x),$$

and $\tilde{f}$ is convex.

Similarly, if $f : \mathcal{V} \to \mathbb{R}$ is defined on the whole space $\mathcal{V}$, then

$$\operatorname*{argmin}_{x \in C} f(x) = \operatorname*{argmin}_{x \in \mathcal{V}} \left[f(x) + \delta_C(x) \right],$$

and the function on the right hand side is convex.

Affine functions

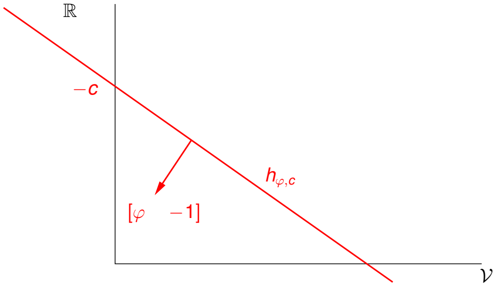
Note: If you do not know what the dual space $\mathcal{V}^*$ of a vector space is, then you can substitute $\mathcal{V}$ - we work the Hilbert space $\mathcal{L}^2(\Omega)$, so they are the same.

Definition

Let $\varphi \in \mathcal{V}^*$ and $c \in \mathbb{R}$, then an affine function on $\mathcal{V}$ is given by

$$h_{\varphi,c} : v \mapsto \langle x, \varphi \rangle - c.$$

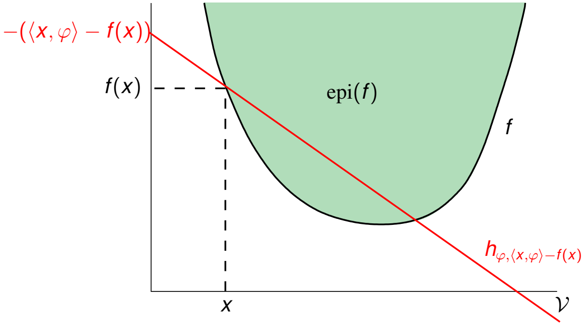
We call φ the slope and c the intercept of $h_{\varphi,c}$.



Affine functions

We would like to find the largest affine function below f. For this, consider for each $x \in \mathcal{V}$ the affine function which passes through $(x, f(x))$:

$$h_{\varphi,c}(x) = f(x) \Leftrightarrow \langle x, \varphi \rangle - c = f(x) \Leftrightarrow c = \langle x, \varphi \rangle - f(x).$$



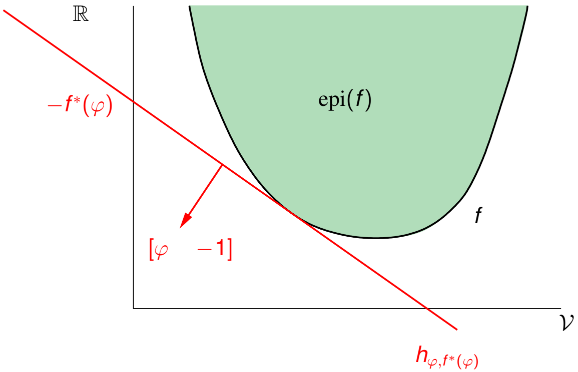
To get the largest affine function below f, we have to pass to the supremum. The intercept of this function is called the conjugate functional of f.

Conjugate functionals

Definition

Let $f \in \text{conv}(\mathcal{V})$. Then the conjugate functional $f^* : \mathcal{V}^* \to \mathbb{R} \cup \{\infty\}$ is defined as

$$f^*(\varphi) := \sup_{x \in \mathcal{V}} [\langle x, \varphi \rangle - f(x)] \,.$$



Second conjugate

The epigraph of f^* consists of all pairs (φ, c) such that $h_{\varphi,c}$ lies below f. It almost completely characterizes f. The reason for the "almost" is that you can recover f only up to closure.

Theorem

Let $f \in \mathrm{conv}(\mathcal{V})$ be closed and $\mathcal{V}$ be reflexive, i.e. $\mathcal{V}^{**} = \mathcal{V}$. Then $f^{**} = f$.

For the proof, note that

$$f(x) = \sup_{h_{\varphi,c} \leq f} h_{\varphi,c}(x) = \sup_{(\varphi,c) \in \mathrm{epi}(f^*)} h_{\varphi,c}(x)$$

$$= \sup_{\varphi \in \mathcal{V}^*} [\langle x, \varphi \rangle - f^*(\varphi)] = f^{**}(x).$$

The first equality is intuitive, but surprisingly difficult to show - it ultimately relies on the theorem of Hahn-Banach.

The subdifferential

Definition

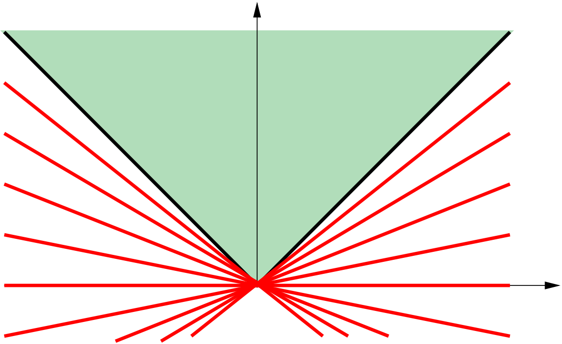
- Let $f \in \operatorname{conv}(\mathcal{V})$. A vector $\varphi \in \mathcal{V}^*$ is called a subgradient of f at $x \in \mathcal{V}$ if
$$f(y) \geq f(x) + \langle y - x, \varphi \rangle \text{ for all } y \in \mathcal{V}.$$
- The set of all subgradients of f at x is called the subdifferential $\partial f(x)$.

Geometrically speaking, φ is a subgradient if the graph of the affine function
$$h(y) = f(x) + \langle y - x, \varphi \rangle$$
lies below the epigraph of f. Note that also $h(x) = f(x)$, so it "touches" the epigraph.

The subdifferential

Example: the subdifferential of $f : x \mapsto |x|$ in 0 is

$$\partial f(0) = [-1, 1].$$



Subdifferential and derivatives

The subdifferential is a generalization of the Fréchet derivative (or the gradient in finite dimension), in the following sense.

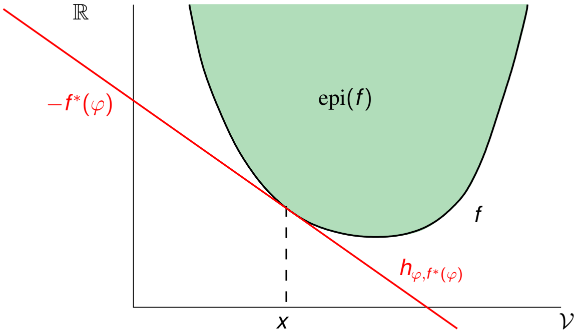
Theorem (subdifferential and Fréchet derivative

Let $f \in \mathrm{conv}(\mathcal{V})$ be Fréchet differentiable at $x \in \mathcal{V}$. Then

$$\partial f(x) = \{df(x)\}.$$

The proof of the theorem is surprisingly involved - it requires to relate the subdifferential to one-sided directional derivatives. We will not explore these relationships in this lecture.

Relationship between subgradient and conjugate



φ is a subgradient at x if and only if the line $h_{\varphi,f^*(\varphi)}$ touches the epigraph at x. In formulas,

$$
\begin{aligned}
& \varphi \in \partial f(x) \\
\Leftrightarrow \quad & h_{\varphi,f^*(\varphi)}(y) = f(x) + \langle y - x, \varphi \rangle \\
\Leftrightarrow \quad & f^*(\varphi) = \langle x, \varphi \rangle - f(x)
\end{aligned}
$$

The subdifferential and duality

The previously seen relationship between subgradients and conjugate functional can be summarized in the following theorem.

Theorem

Let $f \in conv(\mathcal{V})$ and $x \in \mathcal{V}$. Then the following conditions on a vector $\varphi \in \mathcal{V}^*$ are equivalent:

- $\varphi \in \partial f(x)$.
- $x = \operatorname{argmax}_{y \in \mathcal{V}} \left[\langle y, \varphi \rangle - f(y) \right]$.
- $f(x) + f^*(\varphi) = \langle x, \varphi \rangle$.

If furthermore, f is closed, then more conditions can be added to this list:

- $x \in \partial f^*(\varphi)$.
- $\varphi = \operatorname{argmax}_{\psi \in \mathcal{V}^*} \left[\langle x, \psi \rangle - f^*(\psi) \right]$.

Formal proof of the theorem

The equivalences are easy to see.

- Rewriting the subgradient definition, one sees that $\varphi \in \partial f(x)$ means
$$\langle x, \varphi \rangle - f(x) \geq \langle y, \varphi \rangle - f(y) \text{ for all } y \in \mathcal{V}.$$
This implies the first equivalence.
- We have seen the second one on the slide before.
- If f is closed, then $f^{**} = f$, thus we get
$$f^{**}(x) + f^{*}(\varphi) = \langle x, \varphi \rangle .$$
This is equivalent to the last two conditions using the same arguments as above on the conjugate functional.

Variational principle for convex functionals

As a corollary of the previous theorem, we obtain a generalized variational principle for convex functionals. It is a necessary *and sufficient* condition for the (global) extremum.

Corollary (variational principle for convex functionals)

Let $f \in \operatorname{conv}(\mathcal{V})$. Then $\hat{x}$ is a global minimum of f if and only if

$$0 \in \partial f(\hat{x}).$$

Furthermore, if f is closed, then $\hat{x}$ is a global minimum if and only if

$$\hat{x} \in \partial f^*(0),$$

i.e. minimizing a functional is the same as computing the subdifferential of the conjugate functional at 0.

To see this, just set $\varphi = 0$ in the previous theorem.

Moreau's theorem

For the remainder of the lecture, we will assume that the underlying space is a Hilbert space $\mathcal{H}$, for example $\mathcal{L}^2(\Omega)$.

Theorem (geometric Moreau)

Let f be convex and closed on the Hilbert space $\mathcal{H}$, which we identify with its dual. Then for every $z \in \mathcal{H}$ there is a unique decomposition

$$z = \hat{x} + \varphi \text{ with } \varphi \in \partial f(\hat{x}),$$

and the unique $\hat{x}$ in this decomposition can be computed with the proximation

$$\operatorname{prox}_f(z) := \operatorname*{argmin}_{x \in \mathcal{H}} \left\{ \frac{1}{2} \|x - z\|_{\mathcal{H}}^2 + f(x) \right\}.$$

Corollary to Theorem 31.5 in Rockafellar, page 339 (of 423). The actual theorem has somewhat more content, but is very technical and quite hard to digest. The above is the essential consequence.

Proof of Moreau's Theorem

The correctness of the theorem is not too hard to see: if $\hat{x} = \text{prox}_f(z)$, then

$$\begin{aligned} & \hat{x} \in \operatorname*{argmin}_{x \in \mathcal{H}} \left\{ \frac{1}{2} \|x - z\|_{\mathcal{H}}^2 + f(x) \right\} \\ \Leftrightarrow \quad & 0 \in \ \hat{x} - z + \partial f(\hat{x}) \\ \Leftrightarrow \quad & z \in \hat{x} + \partial f(\hat{x}). \end{aligned}$$

Existence and uniqueness of the proximation follows because the functional is closed, strictly convex and coercive.

The geometry of the graph of ∂f

- The map $z \mapsto (\text{prox}_f(z), z - \text{prox}_f(z))$ is a continuous map from $\mathcal{H}$ into the graph of ∂f,

$$\text{graph}(\partial f) := \{(x, \varphi) \; : \; x \in \mathcal{H}, \varphi \in \partial f(x)\} \subset \mathcal{H} \times \mathcal{H},$$

  with continuous inverse $(x, \varphi) \mapsto x + \varphi$.
- The theorem of Moreau now says that this map is one-to one. In particular,

$$\mathcal{H} \simeq \text{graph}(\partial f),$$

  i.e. the sets are homeomorphic.
- In particular, graph(∂f) is always connected.

Fixed points of the proximation operator

Proposition

Let f be closed and convex on the Hilbert space $\mathcal{H}$. Let $\hat{z}$ be a fixed point of the proximation operator prox_f, i.e.

$$\hat{z} = \text{prox}_f(\hat{z}).$$

Then $\hat{z}$ is a minimizer of f. In particular, it also follows that

$$\hat{z} \in (I - \text{prox}_f)^{-1}(0).$$

To proof this, just note that because of Moreau's theorem,

$$\hat{z} \in \text{prox}_f(\hat{z}) + \partial f(\hat{z}) \Leftrightarrow 0 \in \partial f(\hat{z})$$

if $\hat{z}$ is a fixed point.

Subgradient descent

Let $\lambda > 0$, $z \in \mathcal{H}$ and $x = \operatorname{prox}_{\lambda f}(z)$. Then

$$\begin{aligned} z &\in x + \partial \lambda f(x) \\ \Leftrightarrow x &\in z - \lambda \partial f(x). \end{aligned}$$

In particular, we have the following interesting observation:

The proximation operator $\operatorname{prox}_{\lambda f}$ computes an implicit subgradient descent step of step size λ for the functional f.

Implicit here means that the subgradient is not evaluated at the original, but at the new location. This improves stability of the descent. Note that if subgradient descent converges, then it converges to a fixed point $\hat{z}$ of $I - \lambda \partial f$, in particular $\hat{z}$ is a minimizer of the functional f.

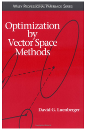
Summary

- Variational calculus deals with functionals on infinite-dimensional vector spaces.
- Minima are characterized by the variational principle, which leads to the Euler-Lagrange equation as a condition for a local minimum.
- The total variation is a powerful regularizer for image processing problems. For binary functions u, it equals the perimeter of the set where $u = 1$.
- Convex optimization deals with finding minima of convex functionals, which can be non-differentiable.
- The generalization of the variational principle for a convex functional is the condition that the subgradient at a minimum is zero.
- Efficient optimization methods rely heavily on the concept of duality. It allows certain useful transformations of convex problems, which will be employed in the next chapter.

References (1)

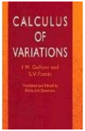
Variational methods

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)

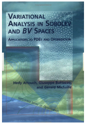
References (2)

Total Variation

Chambolle, Caselles, Novaga, Cremers, Pock
"An Introduction to Total Variation for Image Analysis",
Summer School, Linz, Austria 2006.

- Focused introduction to total variation for image processing applications, plus some basics of convex optimization and the numerics of optimization.
- Available online for free.

Attouch, Buttazzo and Micaille,
"Variational Analysis in Sobolev and BV spaces",
SIAM 2006.

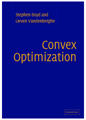


- Exhaustive introduction to variational methods and convex optimization in infinite dimensional spaces, as well as the theory of BV functions.
- Mathematically very advanced, requires solid knowledge of functional analysis.

References (3)

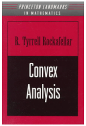
Convex Optimization

Boyd and Vandenberghe,
“Convex Optimization”,
Stanford University Press 2004.



- Excellent recent introduction to convex optimization.
- Reads very well, available online for free.

Rockafellar,
“Convex Analysis”,
Princeton University Press 1970.



- Classical introduction to convex analysis and optimization.
- Somewhat technical and not too easy to read, but very exhaustive.