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
   - An example: Image Denoising
   - The variational principle
   - The Euler-Lagrange equation

2. Total Variation and Co-Area
   - The space $BV(\Omega)$
   - Geometric properties
   - Co-area

3. Convex analysis
   - Convex functionals
   - Duality
   - Conjugate functionals
   - Legendre-Fenchel dual
   - Subdifferential calculus
   - Generalized variational principle
Variational Methods

Introduction
An example: Image Denoising
The variational principle
The Euler-Lagrange equation

Total Variation and Co-Area
The space $BV(\Omega)$
Geometric properties
Co-area

Convex analysis
Convex functionals
Duality
Conjugate functionals
Legendre-Fenchel dual
Subdifferential calculus
Generalized variational principle
Fundamental problems in computer vision

Image labeling problems

Segmentation

Stereo

Optic flow
Fundamental problems in computer vision

3D Reconstruction
Unifying concept: variational approach
Unifying concept: variational approach

Problem solution is the minimizer of an energy functional $E$, 

$$\arg\min_{u \in \mathcal{V}} E(u).$$
Unifying concept: variational approach

Problem solution is the minimizer of an energy functional $E$, 

$$\arg\min_{u \in \mathcal{V}} E(u).$$

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

A greyscale image is a real-valued function

\[ u : \Omega \rightarrow \mathbb{R} \text{ on an open set } \Omega \subset \mathbb{R}^2. \]
Images are functions

A color image is a vector-valued function
\[ u : \Omega \rightarrow \mathbb{R}^3 \] on an open set \( \Omega \subset \mathbb{R}^2 \),
which maps e.g. into RGB color space.
Surfaces are manifolds

Continuous vs. Discrete
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

$$\argmin_{u \in \mathcal{L}^2(\Omega)} \left[ \int_{\Omega} |\nabla u|_2 \, dx + \frac{1}{2\lambda} \int_{\Omega} (u - f)^2 \, dx \right].$$

Note: In Bayesian statistics, this can be interpreted as a MAP estimate for Gaussian noise.
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 $L^2(\Omega)$

- The functional
  $$\|u\|_2 := \left( \int_{\Omega} |u|^2 \, dx \right)^{\frac{1}{2}}$$

  is a norm on $L^2(\Omega)$, with which it becomes a Banach space.

- The norm arises from the inner product
  $$\langle u, v \rangle \mapsto \int_{\Omega} uv \, dx$$

  if you set $\|u\|_2 := \sqrt{\langle u, u \rangle}$. Thus, $L^2(\Omega)$ 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 $L^2(\Omega)$, and convergence, continuity etc. is defined with respect to the above norm.
$\mathbb{R}^n \text{ vs. } L^2(\Omega)$

<table>
<thead>
<tr>
<th>Elements</th>
<th>$\mathbb{R}^n$</th>
<th>$L^2(\Omega)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>finitely many components $x_i, 1 \leq i \leq n$</td>
<td></td>
<td>infinitely many “components” $u(x), x \in \Omega$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inner Product</th>
<th>$(x, y) = \sum_{i=1}^{n} x_i y_i$</th>
<th>$(u, v) = \int_{\Omega} uv , dx$</th>
</tr>
</thead>
</table>

| Norm | $|x|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ | $\|u\|_2 = \left( \int_{\Omega} |u|^2 \, dx \right)^{\frac{1}{2}}$ |
|------|---------------------------------|--------------------------|

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

<table>
<thead>
<tr>
<th>Gradient (Fréchet)</th>
<th>$dE(x) = \nabla E(x)$</th>
<th>$dE(u) =$ ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directional (Gâteaux)</td>
<td>$\delta E(x; h) = \nabla E(x) \cdot h$</td>
<td>$\delta E(u; h) =$ ?</td>
</tr>
<tr>
<td>Condition for minimum</td>
<td>$\nabla E(\hat{x}) = 0$</td>
<td>?</td>
</tr>
</tbody>
</table>
## Definition

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

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

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

- The Gâteaux differential can be thought 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 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 $E : \mathcal{V} \rightarrow \mathbb{R}$, 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 \( E : C^1(\Omega) \rightarrow \mathbb{R} \), and \( E \) be of the form

\[
E(u) = \int_{\Omega} L(u, \nabla u, x) \, dx,
\]

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

\[
\partial_a L(u, \nabla u, x) - \text{div}_x [\nabla_b L(u, \nabla u, x)] = 0,
\]

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

\[
\partial_a L := \frac{\partial L}{\partial a}, \quad \nabla_b L := \left[ \frac{\partial L}{\partial b_1} \ldots \frac{\partial L}{\partial b_n} \right]^T.
\]
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 $u \in \mathcal{L}_{loc}^1(\Omega)$. If

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

for every test function $h \in \mathcal{C}^\infty_c(\Omega)$, then $u = 0$ almost everywhere.
Derivation of Euler-Lagrange equation (1)

Let \( h \in C_c^\infty(\Omega) \) 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

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

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 \( \phi_u \) is the generalization of the gradient, since directional derivatives are computed via the linear map

\[
h \mapsto (\phi_u, h).
\]

The function \( \phi_u \) 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, \( n : \partial \Omega \rightarrow \mathbb{R}^n \) the outer normal of \( \Omega \) and \( p : \mathbb{R}^n \rightarrow \mathbb{R}^n \) a continuously differentiable vector field, defined at least in a neighbourhood of \( \Omega \). Then
\[
\int_{\Omega} \text{div}(p) \, dx = \oint_{\partial \Omega} p \cdot n \, ds.
\]

Corollary: integration by parts
If in addition, \( u : \Omega \rightarrow \mathbb{R} \) is a differentiable scalar function, then
\[
\int_{\Omega} \nabla u \cdot p \, dx = -\int_{\Omega} u \cdot \text{div}(p) \, dx + \oint_{\partial \Omega} up \cdot n \, ds.
\]
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 $p = \nabla_b L(u, \nabla u, x)$, noting $h \big|_{\partial \Omega} = 0$, we get

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

This is the desired expression, from which we can directly see the definition of $\phi_u$. 
Euler-Lagrange equation for the ROF functional (1)

- The ROF functional

$$\int_{\Omega} |\nabla u|_2 + \frac{1}{2\lambda} (u - f)^2 \, dx$$

is of the form

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

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 the Euler-Lagrange equation for an approximated $L_\epsilon$ for a regularization parameter $\epsilon > 0$:

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

$$= : |b|_\epsilon$$
Euler-Lagrange equation for the ROF functional (2)

- The approximation $L_\epsilon$ 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

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

$$= \phi_u$$
Open questions

• 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 \( L^2(\Omega) \). 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

1. Variational Methods
   - Introduction
   - An example: Image Denoising
   - The variational principle
   - The Euler-Lagrange equation

2. Total Variation and Co-Area
   - The space $BV(\Omega)$
   - Geometric properties
   - Co-area

3. Convex analysis
   - Convex functionals
   - Duality
   - Conjugate functionals
   - Legendre-Fenchel dual
   - Subdifferential calculus
   - Generalized variational principle
Definition of the total variation

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

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

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

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

Let $u \in C^1(\Omega)$. Then $J(u) = \int_{\Omega} |\nabla u|_2^2 \, dx$, where $|\cdot|_2$ denotes the Euclidean norm.

To see this, first observe that because of Gauss' theorem,

$$
-\int_{\Omega} u \cdot \text{div}(\xi) \, dx = \int_{\Omega} \nabla u \cdot \xi \, dx \leq \|\xi\|_{\infty} \int_{\Omega} |\nabla u|_2^2 \, dx.
$$

Equality holds for the vector field $\bar{\xi} \in L^1(\Omega, \mathbb{R}^n)$ defined as follows:

$$
\bar{\xi}(x) := \begin{cases} 
\frac{\nabla u(x)}{|\nabla u(x)|_2} & \text{if } |\nabla u(x)|_2 \neq 0 \\
0 & \text{otherwise},
\end{cases}
$$

since $\bar{\xi}$ can be approximated by a sequence $(\xi_n) \subset C^1_c(\Omega)$, and

$$
\int_{\Omega} |\nabla u|_2^2 \, dx = \int_{\Omega} \nabla u \cdot \bar{\xi} \, dx \leftarrow \int_{\Omega} \nabla u \cdot \xi_n \, dx = -\int_{\Omega} u \cdot \text{div}(\xi_n) \, dx.
$$
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 $L^2(\Omega)$.
- $J$ is lower semi-continuous on $L^2(\Omega)$, 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.
We now compute the TV of the characteristic function of a “sufficiently nice” set $U \subset \Omega$, with a $C^1$-boundary.

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

$$- \int_{\Omega} 1_U \cdot \text{div}(\xi) \, dx = - \int_{U} \text{div}(\xi) \, dx$$

$$= \int_{\partial U} n \cdot \xi \, ds \quad \text{(Gauss’ theorem)}$$

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

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

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

---

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 $C^1$-boundary.**

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

$$- \int_{\Omega} 1_U \cdot \text{div}(\xi) \, dx = - \int_{U} \text{div}(\xi) \, dx$$

$$= \int_{\partial U} n \cdot \xi \, ds \quad \text{(Gauss’ theorem)}$$

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

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

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

---

Total Variation and Co-Area

Geometric properties

D. Cremers, B. Goldlücke, T. Pock

ECCV 2010 Tutorial

Variational Methods in Computer Vision
The perimeter of a set

\[ \partial U \]

\[ \{ 1_U = 0 \} \]

\[ U = \{ 1_U = 1 \} \]

In summary, we found that

\[ J(1_U) = \mathcal{H}^{n-1}(\partial U). \]

This motivates the following definition.

**Definition**

Let \( U \subset \Omega \). Then the perimeter of \( U \) in \( \Omega \) is defined as

\[ \text{Per}(U, \Omega) = J(1_U). \]

Note that in view of the above,

\[ \text{Per}(U, \Omega) = J(1_U) = \mathcal{H}^{n-1}(\partial U). \]
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 \( u \in BV(\Omega) \). Then

\[
J(u) = \int_{-\infty}^{\infty} J(1_{\{u \leq t\}}) \, dt = \int_{-\infty}^{\infty} \text{Per}(\{u \leq t\}, \Omega) \, dt.
\]
Variational Methods
Introduction
An example: Image Denoising
The variational principle
The Euler-Lagrange equation

Total Variation and Co-Area
The space $BV(\Omega)$
Geometric properties
Co-area

Convex analysis
Convex functionals
Duality
Conjugate functionals
Legendre-Fenchel dual
Subdifferential calculus
Generalized variational principle
The epigraph of a functional

**Definition**

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

$$\text{epi}(f) := \{(x, \mu) : x \in \mathcal{V} \text{ and } \mu \geq f(x)\}.$$
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 $\text{epi}(f)$ is a convex set.
- The set of all proper and convex functionals on $\mathcal{V}$ is denoted $\text{conv}(\mathcal{V})$.

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.
Convex versus non-convex energies

non-convex energy

convex energy
Convex versus non-convex energies

Convex energies can be globally minimized - for non-convex energies this is usually impossible.
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 lower level sets.
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 $\text{epi}(f)$ 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 $f : \mathcal{V} \rightarrow \mathbb{R} \cup \{\infty\}$ 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 lower semi-continuous on $\mathbb{R}$. The infimum 0 is not attained.
- The functional of the ROF model is closed and convex. It is also coercive on $L^2(\Omega)$: from the inverse triangle inequality,

$$\|u - f\|_2 \leq \|u\|_2 - \|f\|_2.$$

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 $f \in L^2(\Omega)$. 
The study of duality can be thought of as a generalization of Hilbert space theory to spaces without an inner product. The idea is to replace the inner product by a **pairing** of the space with its dual.

**Definition**

- The (topological) **dual space** of a topological vector space $\mathcal{V}$ is the vector space of continuous linear functionals on $\mathcal{V}$,

$$\mathcal{V}^* := \{ \varphi : \mathcal{V} \to \mathbb{R} \text{ linear and continuous} \}.$$  

- The dual pairing of a vector $u \in \mathcal{V}$ and $\varphi \in \mathcal{V}^*$ is the bilinear map

$$\langle u, \varphi \rangle := \varphi(u).$$

Example: the dual space of the $n$-dimensional column vectors $\mathbb{R}^{n \times 1}$ is the space $\mathbb{R}^{1 \times n}$ of $n$-dimensional row vectors. A row $a$ represents a linear functional: if $\mathbf{v}$ is a column, then $a(\mathbf{v})$ is given by the matrix product $a \mathbf{v}$. The functional $\mathbf{v} \mapsto a(\mathbf{v})$ is continuous in the usual topology of $\mathbb{R}^n$. 
Dual of a Hilbert space

The dual of a Hilbert space is particularly simple: It is just the space itself.

Riesz representation theorem

- Let \( H \) be a Hilbert space and \( \varphi \in H^* \). Then there exists a \( v \in H \) such that
  \[
  \langle u, \varphi \rangle = (u, v) \text{ for all } u \in H.
  \]
- Furthermore, for any \( v \in H \),
  \[
  u \mapsto (u, v)
  \]
  is a continuous linear form on \( H \).
- In particular, \( H^* \) can be identified with \( H \), and the dual pairing is given by the inner product.

In this lecture, all our base spaces are Hilbert spaces, so we could live without the notion of the dual space. However, in order to familiarize you with the concept and for the sake of more generality, I will strictly distinguish between the base space and the dual.
Lagrange duality can be used to transform constrained minimization problems to unconstrained “saddle point problems” with more variables.

**Theorem (Lagrange duality)**

Let \( f \in \text{conv}(\mathcal{V}) \), \( g : \mathcal{V} \to \mathcal{W} \) be convex. Then (under certain technical assumptions)

\[
\inf_{x \in \mathcal{V}, g(x)=0} \left\{ f(x) \right\} = \sup_{\lambda \in \mathcal{W}^*} \inf_{x \in \mathcal{V}} \left\{ f(x) + \langle g(x), \lambda \rangle \right\}.
\]

For inequality constraints, one gets a constrained saddle point problem with very simple constraints on \( \lambda \):

\[
\inf_{x \in \mathcal{V}, g(x)\leq 0} \left\{ f(x) \right\} = \sup_{\lambda \in \mathcal{W}^*, \lambda \geq 0} \inf_{x \in \mathcal{V}} \left\{ f(x) + \langle g(x), \lambda \rangle \right\}.
\]

The numbers \( \lambda \) are the **Lagrange multipliers** of the problem.

You will learn much more about saddle point problems in the second part of the tutorial.
Affine functions

In this tutorial, we interpret elements of the dual in a very geometrical way as the slope of affine functions.

Definition

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

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

We call $\varphi$ 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 $(x, f(x))$:

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

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** \( f^* : \mathcal{V}^* \rightarrow \mathbb{R} \cup \{\infty\} \) is defined as

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

An immediate consequence of the definition is

**Fenchel’s inequality**

Let \( f \in \text{conv}(\mathcal{V}) \). Then for all \( x \in \mathcal{V} \) and \( \varphi \in \mathcal{V}^* \),

\[
\langle x, \varphi \rangle \leq f(x) + f^*(\varphi).
\]

In the above equation, equality holds if and only if \( \varphi \) belongs to the **subdifferential** \( \partial f(x) \).
Geometric interpretation of the conjugate functional

\[ -f^*(\varphi) \]

\[ \begin{bmatrix} \varphi & -1 \end{bmatrix} \]

\[ h_{\varphi, f^*}(\varphi) \]
The epigraph of $f^*$ consists of all pairs $(\varphi, c)$ 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 $f \in \text{conv}(\mathcal{V})$ be closed and $\mathcal{V}$ be reflexive, i.e. $\mathcal{V}^{**} = \mathcal{V}$. Then $f^{**} = f$.

For the proof, note that

\[
\begin{align*}
    f(x) &= \sup_{h_{\varphi, c} \leq f} h_{\varphi, c}(x) = \sup_{(\varphi, c) \in \text{epi}(f^*)} h_{\varphi, c}(x) \\
    &= \sup_{\varphi \in \mathcal{V}^*} \left[ \langle x, \varphi \rangle - f^*(\varphi) \right] = f^{**}(x).
\end{align*}
\]

The first equality is intuitive, but surprisingly difficult to show.
A concave functional and conjugate of it are obtained by just mirroring a convex one horizontally.

**Definition**

A functional \( g : \mathcal{V} \rightarrow \mathbb{R} \cup \{-\infty\} \) is called **concave** if \(-g\) is convex. The **hypograph** of \( g \) is the set “below the graph”, i.e.

\[
\text{hyp}(g) = \{(x, \mu) \in \mathcal{V} \times \mathbb{R} : \mu \leq g(x)\}.
\]

The conjugate \( g^* : \mathcal{V}^* \rightarrow \mathbb{R} \cup \{-\infty\} \) of a concave functional \( g \) is defined as

\[
g^*(\varphi) := \inf_{x \in \mathcal{V}} [\langle x, \varphi \rangle - g(x)].
\]
Theorem

Let $f$ be convex and $g$ be concave. Then

$$\inf_{x \in V} [f(x) - g(x)] = \max_{\varphi \in V^*} [g^*(\varphi) - f^*(\varphi)].$$

Geometric intuition: The minimum vertical distance between $\text{epi}(f)$ and $\text{hyp}(g)$ equals the maximum vertical distance between parallel separating hyperplanes.
Theorem

Let $f$ be convex and $g$ be concave. Then

$$\inf_{x \in \mathcal{V}} [f(x) - g(x)] = \max_{\varphi \in \mathcal{V}^*} [g^*(\varphi) - f^*(\varphi)].$$

Geometric intuition: The minimum vertical distance between $\text{epi}(f)$ and $\text{hyp}(g)$ equals the maximum vertical distance between parallel separating hyperplanes.
The subdifferential

Definition

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

Geometrically speaking, \( \varphi \) 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

$$\partial f(0) = [-1, 1].$$
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 \( f \in \text{conv}(\mathcal{V}) \) 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

Here, we can see the equivalence

\[ \varphi \in \partial f(x) \]
\[ \iff h_{\varphi, f^*(\varphi)}(y) = f(x) + \langle y - x, \varphi \rangle \]
\[ \iff f(x) = \langle x, \varphi \rangle - f^*(\varphi) \]
The subdifferential and duality

The main properties of the subdifferential are summarized in the following theorem. It relates the subdifferential to certain optimization problems, as well as to conjugate functionals.

**Theorem**

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

- \( \varphi \in \partial f(x) \).
- \( x = \arg \max_{y \in \mathcal{V}} [\langle y, \varphi \rangle - f(y)] \).
- \( f(x) + f^*(\varphi) = \langle x, \varphi \rangle \).

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

- \( x \in \partial f^*(\varphi) \).
- \( \varphi = \arg \max_{\psi \in \mathcal{V}^*} [\langle x, \psi \rangle - f^*(\psi)] \).
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. Since the supremum over all $y \in \mathcal{V}$ on the right hand side is $f^*(\varphi)$, we get the second together with the Fenchel inequality.

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.

You should draw pictures to visualize the above - draw the graphs of $f$ and $y \mapsto \langle y, \varphi \rangle - f(y)$, for example, and it should become clear what happens.
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 \( f \in \text{conv}(\mathcal{V}) \). 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.
• **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 special class of functionals.

• 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.
Variational methods


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


- 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)
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.
Convex Optimization


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


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