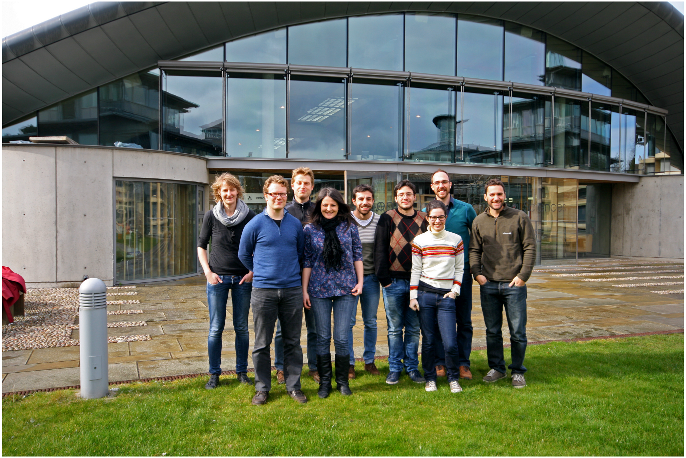
Variational image restoration and segmentation

Jan Lellmann

Cambridge Image Analysis
Department for Applied Mathematics and Theoretical Physics
Cambridge University

Granada, May 2015

Cambridge Image Analysis (CIA)



1 PI, 5 PostDocs, 6 PhD, 3 Master

www.damtp.cam.ac.uk/research/cia/

www.images.group.cam.ac.uk

Overview

Variational methods in image processing

Model

Model

Model

Problem formulation

Given data I, find the (image-)information u so that

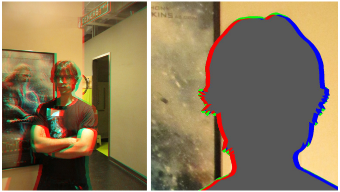
$$I = T(u) + n,$$

where T is the "forward model" describing how the measurements I are generated from u, and n is a random variable modelling the noise.

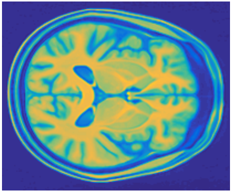
Difficulties

In many applications, reconstructing u from I is

- not unique (T "forgets" data),
- not stable (small errors in $I \to$ large error in u)
- not deterministic due to the random noise n

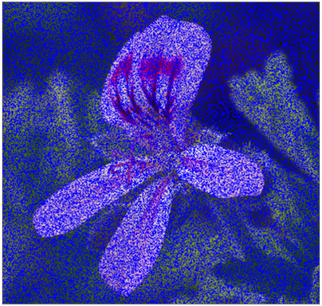
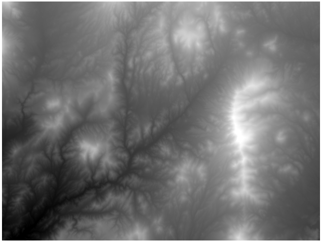
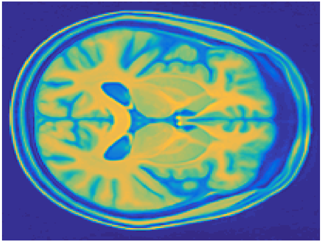


R. Hocking



J. Acosta-Cabronero

What is a "typical" image?

Prior knowledge

Variational method

We reconstruct the (image-) information u from the data I by minimizing an energy

$$\min_u \left\{ \underbrace{D(T(u);I)}_{\substack{\text{data term, compatibility} \\ \text{with measurements } f}} + \underbrace{R(u)}_{\substack{\text{regularizer, prior knowledge} \\ \text{(problem specific)}}} \right\}$$

Advantages:

- Intuitive modeling by specifying properties of desired output
- Often statistical motivation, e.g. *Maximum A Posteriori*-estimate
- Modularity and reusability of individual components

$$R(u) = \int_\Omega \|Du\| \; \longrightarrow$$

Strategies for building regularisers

Trade-offs

model complexity vs. **tractability/computability**

local minimizers vs. **global minimizers**

Top-down approach

- Difficult physical/biological models
- **Advantages:**
  - very specific
  - model parameters contain addition information
- **Disadvantages:**
  - Optimization is difficult

Bottom-up approach

- Combine simple, well-understood components with **adaptivity** and **relaxation**
- **Advantages:**
  - mathematical analysis
  - global minimization
- **Disadvantages:**
  - much less specific

Strategies for building regularisers

Trade-offs

model complexity vs. **tractability/computability**

local minimizers vs. **global minimizers**

Top-down approach

- Difficult physical/biological models
- **Advantages:**
  - very specific
  - model parameters contain addition information
- **Disadvantages:**
  - Optimization is difficult

Bottom-up approach

- Combine simple, well-understood components with **adaptivity** and **relaxation**
- **Advantages:**
  - mathematical analysis
  - global minimization
- **Disadvantages:**
  - much less specific

Overview

Convex Optimization

Extended real valued-functions

In the literature, optimization problems are commonly formulated using an *objective function* $f_0 : \mathbb{R}^n \to \mathbb{R}$ and *constraint functions* $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$, e.g.,

$$\min_{x \in \mathbb{R}^n} f_0(x) \quad s.t. \quad x \in C,$$
$$C = \{x \in \mathbb{R}^n | f_i(x) \leqslant 0, i = 1, \ldots, m\}.$$

By allowing $+\infty$ as the value of the objective function we can rewrite this in a very compact form:

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$, with the definition $x \notin C \Leftrightarrow f(x) = +\infty$.

Extended real-valued calculus

Definition (extended real line)

We define $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$ with the rules:

1. $\infty + c = \infty, \quad -\infty + c = -\infty$ for all $c \in \mathbb{R}$,
2. $0 \cdot \infty = 0, \quad 0 \cdot (-\infty) = 0$,
3. $\inf \mathbb{R} = \sup \emptyset = -\infty, \quad \inf \emptyset = \sup \mathbb{R} = +\infty$.
4. $+\infty - \infty = -\infty + \infty = +\infty$ (sometimes; careful: $-\infty = \lambda(\infty - \infty) \neq \lambda\infty - \lambda\infty = \infty$ if $\lambda < 0$)

Definition (indicator function)

For $C \subseteq \mathbb{R}^n$, denote

$$\delta_C : \mathbb{R}^n \to \bar{\mathbb{R}}, \qquad \delta_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

Constrained minimization

Example (constrained minimization via addition of indicator function)

Assume $f : \mathbb{R}^n \to \mathbb{R}$, $C \subseteq \mathbb{R}^n$, $C \neq \emptyset$. Then

$$x' \text{ minimizes } f \text{ over } C \quad \Leftrightarrow \quad x' \text{ minimizes } f + \delta_C \text{ over } \mathbb{R}^n.$$

Argmin, domain, proper

Definition (argmin, effective domain, proper)

For $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, denote

1. $\operatorname{dom} f := \{x \in \mathbb{R}^n | f(x) < +\infty\}$
2. $\arg\min f := \begin{cases} \emptyset, & f \equiv +\infty, \\ \{x \in \mathbb{R}^n | f(x) = \inf f\}, & f < +\infty. \end{cases}$ (set of minimizers/optimal solutions)
3. f is "proper" $:\Leftrightarrow \operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty \forall x \in \mathbb{R}^n$ (i.e., $f \neq +\infty$ and $f > -\infty$).

Convexity

Definition (convex sets and functions)

1. $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is "convex" $:\Leftrightarrow$

$$f((1-\tau)x + \tau y) \leqslant (1-\tau)f(x) + \tau f(y) \quad \forall x, y \in \mathbb{R}^n, \tau \in (0,1).$$

2. $C \subseteq \mathbb{R}^n$ is "convex" $:\Leftrightarrow \delta_C$ is convex$\Leftrightarrow (1-\tau)x + \tau y \in C \quad \forall x, y \in C, \tau \in (0,1)$.
3. $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is "strictly convex" $:\Leftrightarrow f$ convex and the inequality holds strictly for all $x \neq y$ with $f(x), f(y) \in \mathbb{R}$ and for all $\tau \in (0,1)$.

Global optimality

Theorem (global optimality)

Assume $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *is convex. Then*

1. $\arg\min f$ *is convex.*
2. x *is a local minimizer of* $f \Rightarrow x$ *is a global minimizer of* f.
3. f *strictly convex and proper* $\Rightarrow f$ *has at most one global minimizer.*

Example

1. $\mathbb{R}^n$ is convex,
2. $\{x \in \mathbb{R}^n | x > 0\}$ is convex,
3. $\{x \in \mathbb{R}^n | \|x\|_2 \leqslant 1\}$ is convex,
4. $\{x \in \mathbb{R}^n | \|x\|_2 \leqslant 1, x \neq 0\}$ is *not* convex,
5. the half-spaces $\{x | a^\top x + b \geqslant 0\}$ are convex,
6. $f(x) = a^\top x + b$ is convex (inequality holds as an equality) but *not* strictly convex,
7. $f(x) = \|x\|_2^2$ is strictly convex,
8. $f(x) = \|x\|_2$ is convex but *not* strictly convex.

Convex functions

Theorem (derivative tests)

Assume $C \subseteq \mathbb{R}^n$ is open and convex, and $f : C \to \mathbb{R}$ is differentiable. Then the following conditions are equivalent:

1. *f is [strictly] convex,*
2. *$f(x) + \langle y - x, \nabla f(x) \rangle \leqslant f(y)$ for all $x, y \in C$ [and $< f(y)$ if $x \neq y$],*
3. *$\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.*

If $\nabla^2 f$ is positive definite, then f is strictly convex (but not the other way).

Convexity

Proposition (operations that preserve convexity)

Let $\mathcal{I}$ be an arbitrary index set. Then

1. $f_i, i \in \mathcal{I}$ *convex* $\Rightarrow f(x) := \sup_{i \in \mathcal{I}} f_i(x)$ *is convex,*
2. $f_i, i \in \mathcal{I}$ *strictly convex,* $\mathcal{I}$ *finite* $\Rightarrow f(x) := \sup_{i \in \mathcal{I}} f_i(x)$ *strictly convex,*

Convex calculus

Proposition

1. *(nonnegative linear combination) Assume* $f_1, \dots, f_m : \mathbb{R}^n \to \bar{\mathbb{R}}$ *are convex,* $\lambda_1, \dots, \lambda_m \geqslant 0$. *Then* $f := \sum_{i=1}^m \lambda_i f_i$ *is convex. If at least one of the* f_i *with* $\lambda_i > 0$ *is strictly convex, then* f *is strictly convex.*
2. *(linear composition) Assume* $f : \mathbb{R}^m \to \bar{\mathbb{R}}$ *is convex,* $A \in \mathbb{R}^{m \times n}$*, and* $b \in \mathbb{R}^m$. *Then*

$$g(x) \quad := \quad f(Ax + b)$$

is convex.

Overview

How to build a regularizer

Tikhonov regularization

- A standard assumption is that images are smooth in some way, i.e., they do not oscillate too much.
- This means that the gradient will generally be small, except at boundaries
- $\rightarrow$ penalize the norm of the gradient!

Example (Tikhonov regularization for denoising)

Given $I : \Omega \rightarrow \mathbb{R}$, find

$$\min_{u:\Omega\rightarrow\mathbb{R}} f(u) := \frac{1}{2}\int_\Omega \|u - I\|^2 dx + \lambda \int_\Omega \|\nabla u\|^2 dx.$$

Tikhonov regularization – numerical solution

Example (Tikhonov regularization for denoising)

Given $I : \Omega \to \mathbb{R}$, find

$$\min_{u:\Omega\to\mathbb{R}} f(u) := \frac{1}{2}\int_\Omega \|u - I\|^2 dx + \lambda \int_\Omega \|\nabla u\|^2 dx.$$

Example (Tikhonov discretized problem)

Given $I \in \mathbb{R}^n$, find

$$\begin{aligned}\min_{u\in\mathbb{R}^n} f(u) \quad &:= \quad \frac{1}{2}\sum_{i=1}^{n}(u_i - I_i)^2 + \lambda \sum_{i=1}^{n} \|G_i u\|_2^2 \\ &= \quad \frac{1}{2}\|u - I\|_2^2 + \lambda \|Gu\|_2^2\end{aligned}$$

Tikhonov regularization – numerical solution

Example (Tikhonov discretized problem)

$$\min_{u \in \mathbb{R}^n} f(u) = \frac{1}{2}\|u - I\|_2^2 + \lambda \|Gu\|_2^2$$

Solving the Tikhonov problem

The discretized energy f is convex (and even strictly convex). Therefore any local minimizer is a global minimizer. We know from Fermat's principle that for differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, any local minimizer $u^* \in \mathbb{R}^n$ of f must satisfy

$$\nabla f(u^*) = 0.$$

This leads to a sparse linear equation system which can be solved fast:

$$u - I + \lambda G^\top G u = 0 \quad \Rightarrow \quad (Id + \lambda G^\top G)u = I.$$

Tikhonov regularization

Tikhonov regularization

Total variation

- Tikhonov regularization removes edges – why? The quadratic regularizer makes continuous small change cheaper than sudden big changes with the same height.
- $\rightarrow$ Change the exponent!

Example (TV regularization for denoising, Rudin-Osher-Fatemi)

Given $I : \Omega \rightarrow \mathbb{R}$, find

$$\min_{u:\Omega\rightarrow\mathbb{R}} f(u) := \frac{1}{2}\int_\Omega \|u - I\|^2 dx + \lambda \underbrace{\int_\Omega \|Du\|}_{=:TV(u)}.$$

Remark: For a correct definition in the function space, the gradient ∇u has been replaced by a "distributional gradient" Du, but for the discretized problem it generally does not make a difference.

TV regularization

TV regularization

TV regularization

Total variation

Motivation: We would like to define something like

$$f(u) \;=\; \int_\Omega \|\nabla u(x)\|_2 dx, \tag{1}$$

but this needs u to be differentiable. How to do it for non-differentiable u?

Definition

For $u:\Omega\to\mathbb{R}$, the *total variation (TV)* of u is defined as

$$\mathrm{TV}(u) \;:=\; \sup_{v\in C_c^1(\Omega,\mathbb{R}^n),\|v\|_\infty\leqslant 1} \int_\Omega \langle u, \mathrm{Div}\, v\rangle dx, \tag{2}$$

where $\mathrm{Div}\, v = \partial_{x_1} v_1 + \ldots + \partial_{x_n} v_n$, and for $v\in C_c^1(\Omega,\mathbb{R}^n)$,

$$\|v\|_\infty \;=\; \sup_{x\in\Omega} \|v(x)\|_2.$$

Geometric properties

For characteristic functions, the total variation is just the *length of the boundary* of the underlying set (compare the meaning in 1D):

Proposition

Assume $A \subset \Omega$ *is a set so that its boundary is sufficiently smooth and satisfies* $Len(\Omega \cap \partial A) < \infty$. *Define*

$$1_A(x) \quad := \quad \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Then

$$\mathrm{TV}(1_A) \quad = \quad Len(\Omega \cap \partial A).$$

Total variation can therefore be seen as a “geometric” regularizer that penalizes the *length of the jump set*.

Overview

How to solve the discretized problem?

TV is non-smooth

Example (TV regularization for denoising)

$$\min_{u:\Omega\to\mathbb{R}} f(u) := \frac{1}{2}\int_\Omega \|u - I\|^2 dx + \lambda \int_\Omega \|Du\|_2.$$

Example (TV discretized problem)

Given $I \in \mathbb{R}^n$, find

$$\min_{u\in\mathbb{R}^n} f(u) \quad := \quad \frac{1}{2}\sum_{i=1}^{n}(u_i - I_i)^2 + \lambda \sum_{i=1}^{n} \|G_i u\|_2$$

The energy f is convex but not differentiable! This means we cannot simply use Fermat's principle but have to use specialized non-smooth convex optimization methods based on *primal-dual* or *conic programming* formulations.

Linear programs

Many solvers for non-differentiable convex problems require the problem to be rewritten in a standard form. A classical form ist the Linear Program (LP):

Definition (Linear Program, LP)

For $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, solve

$$\begin{aligned} &\inf_{x \in \mathbb{R}^n} c^\top x \\ &\text{s.t. } Ax \geqslant b, \end{aligned}$$

where the inequality is meant element-wise.

Linear programs

Example

This is surprisingly powerful:

$$\min_x |x_1 - x_2| \quad \text{s.t.} \quad x_1 = -1, x_2 \geqslant 0$$

$$\rightsquigarrow \min_{x,y} y \quad \text{s.t.} \quad y \geqslant |x_1 - x_2|, x_1 \geqslant -1, x_1 \leqslant -1, x_2 \geqslant 0,$$

$$\rightsquigarrow \min_{x,y} y \quad \text{s.t.} \quad y \geqslant x_1 - x_2, y \geqslant x_2 - x_1, x_1 \geqslant -1, -x_1 \geqslant 1, x_2 \geqslant 0,$$

and finally

$$\min_{(x_1,x_2,y)\in\mathbb{R}^3} \quad \underbrace{(0,0,1)}_{c^\top} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix}}_{x} \text{s.t.} \underbrace{\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix}}_{x} \geqslant \underbrace{\begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}}_{b}.$$

Cones

Many image processing problems cannot be written in LP form. However we would still like to keep a similar standard form. The solution is to generalize what "$\geqslant$" means in the $Ax \geqslant b$ constraint.

Definition (cone)

$K \subseteq \mathbb{R}^n$ "cone" $:\Leftrightarrow$

$$0 \in K, \qquad \lambda x \in K \quad \forall x \in K, \lambda \geqslant 0.$$

Note that cones can also be *nonconvex*, such as the cone $K = (\mathbb{R}_{\geqslant 0} \times \{0\}) \cup (\{0\} \times \mathbb{R}_{\geqslant 0})$.

Generalized inequalities

Proposition (generalized inequalities)

For a closed convex cone $K \subseteq \mathbb{R}^n$ we define the "generalized inequality"

$$x \geqslant_K y \quad :\Leftrightarrow \quad x - y \in K.$$

This "behaves" like the usual $\geqslant$ relation:

1. $x \geqslant_K x$ *(reflexivity),*
2. $x \geqslant_K y$, $y \geqslant_K z \Rightarrow x \geqslant_K z$ *(transitivity),*
3. $x \geqslant_K y \Rightarrow -y \geqslant_K -x$ *and* $x \geqslant_K y, \lambda \geqslant 0 \Rightarrow \lambda x \geqslant_K \lambda y$,
4. $x \geqslant_K y, x' \geqslant_K y' \Rightarrow x + x' \geqslant_K y + y'$,
5. *If* $x^k \to x$ *and* $y^k \to y$ *with* $x^k \geqslant_K y^k$ *for all* $k \in \mathbb{N}$, *then* $x \geqslant_K y$.

If "$\geqslant$" is a relation on $\mathbb{R}^n$ satisfying 1.-5., then it can be represented as $\geqslant_K$ for a closed convex cone.

Conic programs

Definition (conic program)

For any closed, convex cone $K \subseteq \mathbb{R}^m$, a matrix $A \in \mathbb{R}^{m \times n}$ and vectors $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, we define the "conic program" or "conic problem" (CP)

$$\inf_x c^\top x \quad \text{s.t.} \quad Ax \geqslant_K b.$$

Many commercial and free "out of the box" solvers require the problem to be reformulated as a conic problem or similar standard form!

Conic programs

Example (standard cone)

The "standard cone"

$$K_n^{\mathrm{LP}} \quad := \quad \{x \in \mathbb{R}^n | x_1, \ldots, x_n \geqslant 0\}$$

is a closed, convex cone. The associated conic program is the "linear program" (LP)

$$\begin{aligned} &\inf_x c^\top x \\ &\text{s.t. } Ax \geqslant b. \end{aligned}$$

Second-order cone

Example (second-order cone)

The "second-order cone" (also called "Lorentz cone", "ice-cream cone")

$$K_n^{\text{SOCP}} \quad := \quad \left\{ x \in \mathbb{R}^n | x_n \geqslant \sqrt{x_1^2 + \ldots + x_{n-1}^2} \right\}$$

is a pointed, closed, convex cone. Conic programs with
$K = K_{n_1}^{\text{SOCP}} \times \ldots \times K_{n_l}^{\text{SOCP}}$ are called "second-order conic programs" (SOCP).

Second-order cone programs

Example

$$\min_x \quad \|x\|_2 \quad \text{s.t.} \quad x_1 + x_2 \geqslant 1.$$

This can be rewritten as a second-order cone program:

$$\begin{aligned} \min_{x,y} \quad & y \\ \text{s.t.} \quad & y \geqslant \|x\|_2, x_1 + x_2 - 1 \geqslant 0. \\ & \Leftrightarrow Id \begin{pmatrix} x \\ y \end{pmatrix} \geqslant_{K_3^{\text{SOCP}}} 0, \quad \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \geqslant_{K_1^{\text{SOCP}}} \begin{pmatrix} 1 \end{pmatrix}. \end{aligned}$$

The LP constraints (inequalities) can also be written as a simple SOCP constraint, but they are usually left in linear form to make notation simpler.

Total Variation as SOCP

- We would like to reformulate the total variation energy

$$\min_{u \in \mathbb{R}^n} f(u) \quad := \quad \frac{1}{2}\sum_{i=1}^{n} |u_i - I_i| + \lambda \sum_{i=1}^{n} \|G_i u\|_2$$

  in conic program form (removing the square in the data term makes it simpler – quadratic data terms require either so-called semidefinite cones or conic programs with quadratic objective).

Total Variation as SOCP

- We would like to reformulate the total variation energy

$$\min_{u \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{n} |u_i - I_i| + \lambda \sum_{i=1}^{n} \|G_i u\|_2$$

Total Variation as SOCP

- Step 1: introduce auxiliary variables for the data term:

$$\begin{aligned} \min_{u\in\mathbb{R}^n, s\in\mathbb{R}^n} \quad & \frac{1}{2}\sum_{i=1}^{n} s_i + \lambda \sum_{i=1}^{n} \|G_i u\|_2 \\ \text{s.t.} \quad & s_i \geqslant |u_i - I_i|, \quad i = 1,\ldots,n. \end{aligned}$$

Total Variation as SOCP

- Step 2: introduce auxiliary variables for the regularizer:

$$\begin{aligned} \min_{u\in\mathbb{R}^n, s\in\mathbb{R}^n, t\in\mathbb{R}^n} \quad & \frac{1}{2}\sum_{i=1}^{n} s_i + \lambda \sum_{i=1}^{n} t_i \\ \text{s.t.} \quad & s_i \geqslant |u_i - I_i|, \quad i = 1, \dots, n, \\ & t_i \geqslant \|G_i u\|_2, \quad i = 1, \dots, n. \end{aligned}$$

Total Variation as SOCP

- Step 2: introduce auxiliary variables for the regularizer:

$$
\begin{aligned}
\min_{u\in\mathbb{R}^n, s\in\mathbb{R}^n, t\in\mathbb{R}^n} \quad & \frac{1}{2}\sum_{i=1}^n s_i + \lambda \sum_{i=1}^n t_i \\
\text{s.t.} \quad & s_i \geqslant |u_i - I_i|, \quad i = 1, \dots, n, \\
& t_i \geqslant \|G_i u\|_2, \quad i = 1, \dots, n.
\end{aligned}
$$

The objective function is now linear, and the constraints have second-order conic (SOCP) form. We may have to introduce more variables to bring the problem into the specific form that the solver requires.

Higher-order Total Variation

- Total variation keeps edges but introduces "stair-casing" artifacts on continuous gradients
- Idea: Penalizing the first derivative keeps the function piecewise constant. Penalizing the second derivatives should keep the function piecewise linear:

$$\min_{u:\Omega\to\mathbb{R}} f(u) := \frac{1}{2}\int_\Omega \|u - I\|^2 dx + \lambda \underbrace{\int_\Omega \|D^2 u\|_2}_{=:TV^2(u)}$$

  where D^2 is the (generalized) Hessian of u.
- We can also use third- or even higher-order derivatives $D^k u$.
- Problem: we cannot have jumps again!

Infimal convolution

Definition (Infimal convolution)

For functions $f_1, \ldots, f_n : X \to \bar{\mathbb{R}}$ for any set X, we define the infimal convolution/inf-convolution $(f_1 \Box \cdots \Box f_n) : X \to \bar{\mathbb{R}}$ as

$$(f_1 \Box \cdots \Box f_k)(u) = \inf_{z^1, \ldots, z^k, z^1 + \ldots + z^k = u} (f_1(z^1) + \ldots + f_k(z^k)).$$

Infimal convolution

Definition (Infimal convolution)

For functions $f_1, \ldots, f_n : X \to \bar{\mathbb{R}}$ for any set X, we define the infimal convolution/inf-convolution $(f_1 \Box \cdots \Box f_n) : X \to \bar{\mathbb{R}}$ as

$$(f_1 \Box \cdots \Box f_k)(u) = \inf_{z^1,\ldots,z^k, z^1+\ldots+z^k=u} (f_1(z^1) + \ldots + f_k(z^k)).$$

Example (ROF is inf-convolution)

$$\min_{u:\Omega\to\mathbb{R}} \frac{1}{2} \int_\Omega \|u - I\|^2 dx + \lambda \int_\Omega \|Du\|_2 dx.$$

Infimal convolution

Definition (Infimal convolution)

For functions $f_1, \ldots, f_n : X \to \bar{\mathbb{R}}$ for any set X, we define the infimal convolution/inf-convolution $(f_1 \Box \cdots \Box f_n) : X \to \bar{\mathbb{R}}$ as

$$(f_1 \Box \cdots \Box f_k)(u) = \inf_{z^1,\ldots,z^k, z^1+\ldots+z^k=u} (f_1(z^1) + \ldots + f_k(z^k)).$$

Example (ROF is inf-convolution)

$$\min_{u,w:\Omega\to\mathbb{R},\ u+w=I} \underbrace{\frac{1}{2}\int_\Omega \|w\|^2 dx}_{f_1(w)} + \underbrace{\lambda \int_\Omega \|Du\|_2 dx}_{f_2(u)}.$$

$\to$ "Cartoon-texture decomposition"

Infimal convolution

- We can use infimal convolution to combine first- and second-order regularizers!

Example ($TV - TV^2$ regularization)

$$\inf_{u,v,w,u+v+w=I} \left\{ \frac{1}{2}\|w\|_2^2 + \lambda\,\mathrm{TV}(u) + \mu\,\mathrm{TV}^2(v) \right\}.$$

- This naturally splits the image into parts
  - With small "Gaussian" energy (w)
  - With few nonzero gradient/piecewise constant (u)
  - With few nonzero second derivatives/piecewise affine, but without jumps (v)

Total Generalized Variation

- We can also split the gradient instead:

Example (Total Generalized Variation, cascading formulation)

$$\inf_{u,v,w,v+w=Du} \left\{ \frac{1}{2}\|u - I\|_2^2 + \lambda \int_\Omega |v| + \mu \int_\Omega \|\mathcal{E}w\|_2 \right\}.$$

Classically $\mathcal{E}$ is the symmetrized gradient, $\frac{1}{2}Dw + \frac{1}{2}(Dw)^\top$.

Overview

Multichannel images

Vector-TV

- What if we have $u : \Omega \to \mathbb{R}^m$ instead?
- Natural extension:

$$\min_{u:\Omega\to\mathbb{R}^m} \frac{1}{2}\int_\Omega \|u(x) - I(x)\|_2^2 dx + \lambda \int_\Omega \|Du\|_\#.$$

- Which norm for $\|\cdot\|_\#$? $Du(x) \in \mathbb{R}^{n\times m}$ (in the smooth case) $\to$ need a matrix norm!
- Common choices: For $A = (a^1, \ldots, a^m) \in R^{n\times m}$,

$$\|A\|_\# = \|a^1\|_2 + \cdots + \|a^m\|_2 \qquad \text{channel-by-channel}$$

$$\|A\|_\# = \left(\sum_{i=1}^m \sum_{j=1}^n (a_j^i)^2\right)^{1/2} \qquad \text{Frobenius norm}$$

- Both representable as an SOCP!

Singular value-based norms

- Idea: A symmetric matrix $A \in \mathbb{R}^{n \times n}$ can be written in the form

$$A = P \Sigma P^\top,$$

where P is orthogonal ($PP^\top = P^\top P = Id$) and $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$ is a diagonal matrix with the eigenvalues $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_n$ of A.

- We could use norms derived from the eigenvalues, for example

$$\|A\|_\# = |\sigma_1| + \ldots + |\sigma_n|, \qquad \text{nuclear norm}$$

$$\|A\|_\# = \max_{i=1,\ldots,n} |\sigma_i|, \qquad \text{spectral norm}$$

$$\|A\|_\# = (|\sigma_1|^p + \cdots + |\sigma_n|^p)^{1/p} \qquad \text{Schatten-p-norm}$$

- But: For $u : \Omega \to \mathbb{R}^m$, we usually have $\nabla u(x) \in \mathbb{R}^{n \times m}$, which is neither quadratic nor symmetric.

Singular values

Proposition (Singular Value Decomposition)

Every matrix $A \in \mathbb{R}^{n \times m}$ *has a singular value decomposition (SVD) of the form*

$$A = U \Sigma V^\top,$$

where $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{m \times m}$, *U and V are unitary matrices, i.e.,* $U^\top U = Id_{n \times n}$, $V^\top V = Id_{m \times m}$, *and* $\Sigma \in \mathbb{R}^{n \times m}$ *is a matrix with the unique (!) singular values* $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$ *on the diagonal, and zero everywhere else.*

SVD

Example (SVD I)

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 8 & 4 \end{pmatrix} = U\Sigma V^\top, \quad \Sigma = \begin{pmatrix} 9.88 & 0 \\ 0 & 2.72 \\ 0 & 0 \end{pmatrix},$$

$$U = \begin{pmatrix} 0.20 & 0.40 & -0.89 \\ 0.40 & 0.80 & 0.45 \\ 0.90 & -0.44 & 0 \end{pmatrix}, \; V = \begin{pmatrix} 0.83 & -0.56 \\ 0.56 & 0.83 \end{pmatrix}$$

Example (SVD II)

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 4 & 8 \end{pmatrix} = U\Sigma V^\top, \quad \Sigma = \begin{pmatrix} 10.25 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$U = \begin{pmatrix} 0.22 & -0.97 & -0.11 \\ 0.44 & 0 & 0.90 \\ 0.87 & 0.24 & -0.42 \end{pmatrix}, \; V = \begin{pmatrix} 0.45 & -0.89 \\ 0.89 & 0.45 \end{pmatrix}$$

SVD

Example (SVD II)

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 4 & 8 \end{pmatrix} = U\Sigma V^\top, \quad \Sigma = \begin{pmatrix} 10.25 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$U = \begin{pmatrix} 0.22 & -0.97 & -0.11 \\ 0.44 & 0 & 0.90 \\ 0.87 & 0.24 & -0.42 \end{pmatrix}, \; V = \begin{pmatrix} 0.45 & -0.89 \\ 0.89 & 0.45 \end{pmatrix}$$

- The number of non-zero singular values is the same as the rank of the matrix.
- We can use regularizers based on the singular values, for example the nuclear norm, Schatten p-norms etc.

Nuclear norm

Definition (nuclear norm)

For $A \in \mathbb{R}^{n \times m}$, the nuclear norm is defined as

$$\|A\|_* = |\sigma_1| + \cdots + |\sigma_n|,$$

where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n$ are the singular values of A.

Example (Nuclear norm TV/low rank regularization)

$$\min_{u:\Omega \to \mathbb{R}} \frac{1}{2} \int_\Omega \|u - I\|_2^2 dx + \lambda \int_\Omega \|Du\|_*.$$

How can we get this into a standard form? It is not an SOCP!

Semidefinite cone

Definition

The "positive semidefinite cone"

$$K_n^{\text{SDP}} \quad := \quad \{X \in \mathbb{R}^{n\times n} | X \text{ symmetric positive semidefinite}\}$$

is a closed, convex cone. Conic programs with $K = K_{n_1}^{\text{SDP}} \times \ldots \times K_{n_l}^{\text{SDP}}$ are called "semidefinite programs" (SDP):

$$\begin{array}{rl} \inf_{x\in\mathbb{R}^n} & c^\top x \\ \text{s.t.} & Ax - b \in K. \end{array}$$

Here A is a linear operator $A : \mathbb{R}^n \to \mathbb{R}^{n_1\times n_1} \times \cdots \times \mathbb{R}^{n_l\times n_l}$, and $b \in \mathbb{R}^{n_1\times n_1} \times \cdots \times \mathbb{R}^{n_l\times n_l}$. Often x and c are also written as matrices $X, C \in \mathbb{R}^{n\times n}$ with the inner product $\langle C, X\rangle := \sum_{i,j} C_{ij}X_{ij}$ replacing $c^\top x$.

Nuclear norm as SDP

Proposition

For $M \in \mathbb{R}^{n\times m}$, the nuclear norm can be written in SDP form:

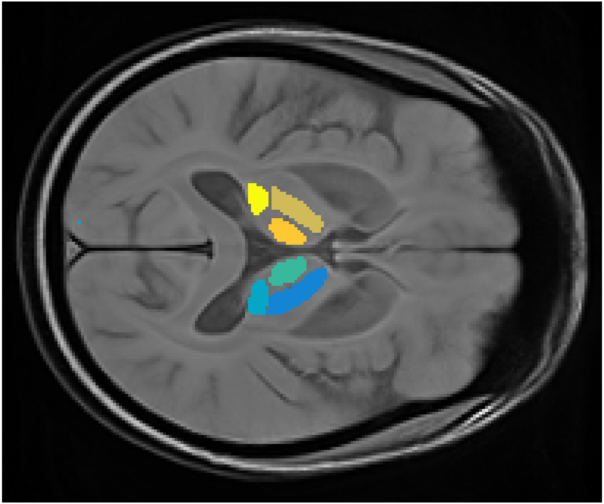
$$\begin{aligned} \|M\|_* &= |\sigma_1| + \cdots + |\sigma_n| \\ &= \min_{B\in\mathbb{R}^{n\times n}, C\in\mathbb{R}^{m\times m}} \frac{1}{2}\mathrm{tr}B + \frac{1}{2}\mathrm{tr}C \\ &\qquad \text{s.t.} \begin{pmatrix} B & M \\ M^\top & C \end{pmatrix} \text{ positive semidefinite.} \end{aligned}$$

This can be proven rigorously using the SVD of A, but as a motivation consider the case $n = 1$, i.e., $M = a \in \mathbb{R}$: Then the semidefiniteness means $b \geqslant 0$, $c \geqslant 0$, and $bc - a^2 \geqslant 0$, i.e., $bc \geqslant a^2$. We compute the minimum of $\frac{1}{2}(b + c)$ subject to $bc \geqslant a^2$. The latter will hold with equality, thus by substitution and optimality conditions $b = c = |a|$.

Segmentation and Relaxation

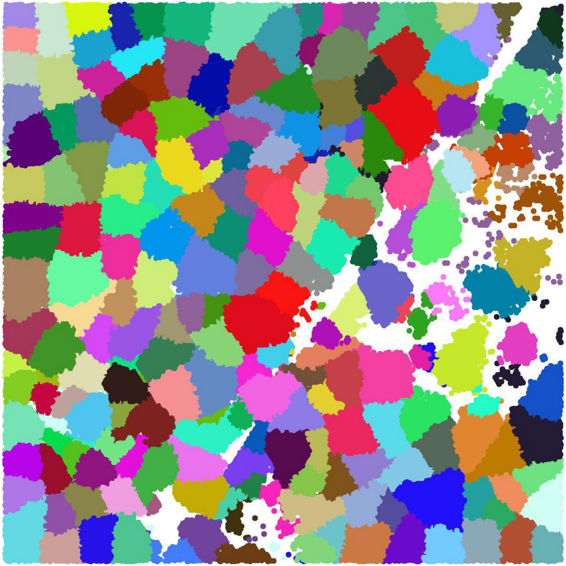
Segmentation

- In many interesting applications the range is discrete:
  In every $x \in \Omega$, a discrete decision has to be made.



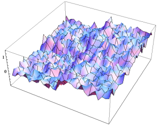
Segmentation

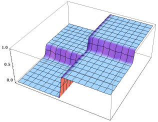
Segmentation

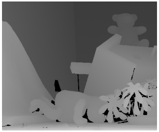


Motivation – Multiclass Labeling

- **Applications:** Segmentation, denoising, 3D reconstruction, depth from stereo, inpainting, photo montage, optical flow,...

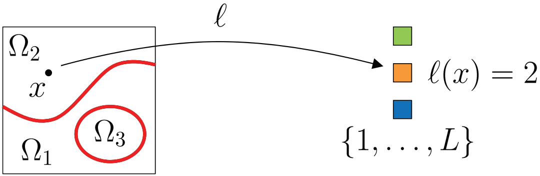




Motivation – Problem

- Finite labeling problem:



- Partition image domain Ω into L regions
- *Discrete* decision at each point in *continuous* domain Ω
- Variational Approach:

$$\min_{\ell:\Omega\to\{1,\dots,L\}} \underbrace{\int_\Omega s(\ell(x),x)dx}_{\text{local data fidelity}} + \underbrace{J(\ell)}_{\text{regularizer}}$$

- Why this form?

Bayesian model

- Assume the domain Ω consists of a finite set of points (pixels) $x \in \Omega$ and we have a local probabilistic model for the observation $I : \Omega \to \mathbb{R}^m$ given the ground-truth (true) segmentation $\ell(x)$,

$$\mathbb{P}(I|\ell) = \prod_{x \in \Omega} \mathbb{P}(I(x)|\ell(x)).$$

- Classic example: Two classes, Gaussian distribution:

$$\begin{aligned}
\mathbb{P}(I(x)|\ell(x) = 1) &= \mathcal{N}(I(x); c_1, \sigma_1^2) \sim e^{-\frac{(I(x)-c_1)^2}{\sigma_1^2}} \\
\mathbb{P}(I(x)|\ell(x) = 2) &= \mathcal{N}(I(x); c_2, \sigma_2^2) \sim e^{-\frac{(I(x)-c_2)^2}{\sigma_2^2}}
\end{aligned}$$

Bayesian model

- Given the observation I, we would like to find the segmentation ℓ that maximizes the "a posteriori probability". Using the Bayes rule:

$$\max_{\ell} \mathbb{P}(\ell|I) = \frac{\mathbb{P}(\ell, I)}{\mathbb{P}(I)} = \frac{\mathbb{P}(I|\ell)\mathbb{P}(\ell)}{\mathbb{P}(I)}.$$

- The $\mathbb{P}(I)$ part is not relevant, and instead of maximizing $\mathbb{P}(\ell|I)$ we can find ℓ by minimizing $-\log \mathbb{P}(\ell|I)$:

$$\min_{\ell} -\log(\mathbb{P}(I|\ell)\mathbb{P}(\ell)) = -\log \mathbb{P}(I|\ell) - \log \mathbb{P}(\ell)$$

$$= \underbrace{\left\{ \sum_{x\in\Omega} -\log \mathbb{P}(I(x)|\ell(x)) \right\}}_{\approx \int_\Omega s(\ell(x),x)dx} \underbrace{-\log \mathbb{P}(\ell)}_{J(\ell)}$$

- For the Gaussian model:

$$s(1,x) = (I(x) - c_1)^2/\sigma_1^2, \quad s(2,x) = (I(x) - c_2)^2/\sigma_2^2.$$

Combinatorial issues

- Approach:

$$\min_{\ell:\Omega\to\{1,\dots,L\}} \underbrace{\int_\Omega s(\ell(x),x)dx}_{\text{local data fidelity}} + \underbrace{J(\ell)}_{\text{regularizer}}$$

- This is a combinatorial problem – optimization is hard!
  - No gradient, Hessian $\to$ no gradient descent, Newton, no simple optimality conditions
  - For n pixels we would have to test all possible L^n assignments
- Idea: Can we extend ("relax") the problem to a larger set of functions and make it convex?

Relaxation example

- Simple combinatorial optimization problem:

$$\min_{x \in \{1,2,3\}} f(x), \quad f(1) = 2, f(2) = 1, f(3) = 3.$$

- We can write this as

$$\min_{x \in \mathbb{R}} g(x), \quad g(x) = \begin{cases} 2, & x = 1, \\ 1, & x = 2 \\ 3, & x = 3 \\ +\infty, & \text{otherwise.} \end{cases}$$

- This g is not convex (and not even finite everywhere)! We would like to find some *convex* function h with $h(x) = g(x)$ if $x \in \{1, 2, 3\}$.
- This is not a unique problem, but if we some choices are better – we do not want to create additional minimizers!

Convex envelope

Definition (Legendre-Fenchel Transform)

Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, then

$$\begin{aligned} f^* &: \mathbb{R}^n \to \bar{\mathbb{R}}, \\ f^*(v) &:= \sup_{x \in \mathbb{R}^n} \{\langle v, x \rangle - f(x)\} \end{aligned}$$

is the "conjugate to f". The mapping $f \mapsto f^*$ is the "Legendre-Fenchel transform". The function $f^{**} = (f^*)^*$ is the "biconjugate" of f.

Theorem (Convex envelope)

Assume $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and assume that the largest convex function g with $g \leqslant f$ is proper. Then the biconjugate f^{**} is the largest convex (lower semi-continuous) function smaller or equal to f.

Relaxation example

- Example:

$$g(x) = \begin{cases} 2, & x = 1, \\ 1, & x = 2 \\ 3, & x = 3 \\ +\infty, & \text{otherwise.} \end{cases}$$

$$g^*(y) \quad = \quad \sup_{x \in \mathbb{R}} \{xy - g(x)\} = \max\{y - 2, 2y - 1, 3y - 3\},$$

For a given slope y, the affine function

$$xy - g^*(y)$$

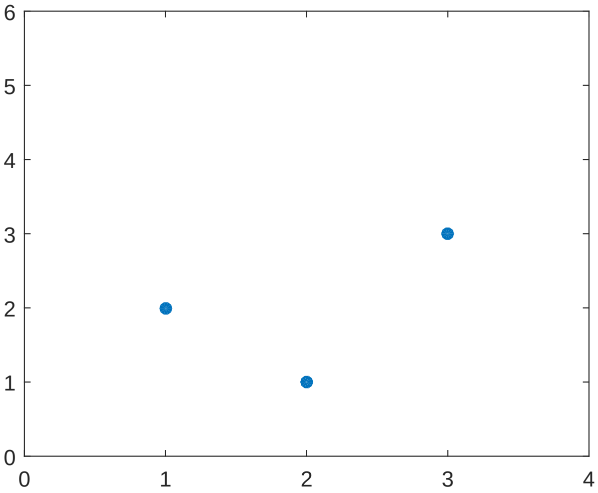
is *below* g and *touches* g from below.

- The biconjugate

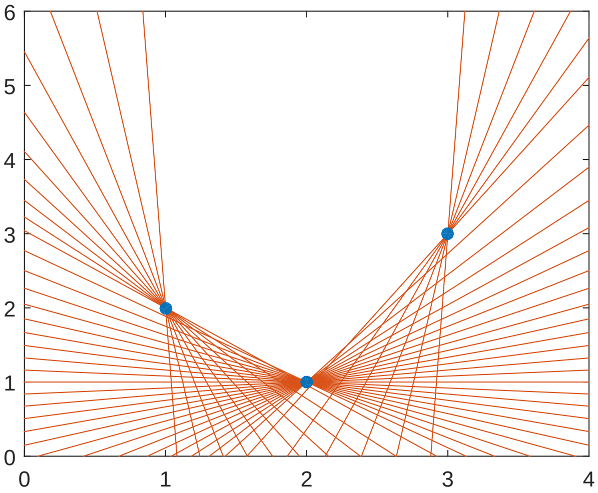
$$g^{**}(x) = \sup_{y \in \mathbb{R}} \{yx - g^*(x)\}$$

is the *pointwise supremum* of all affine functions that touch g from below.

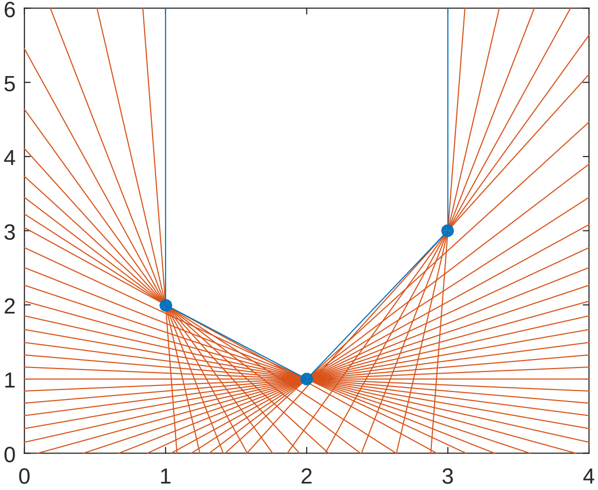
Relaxation



Relaxation

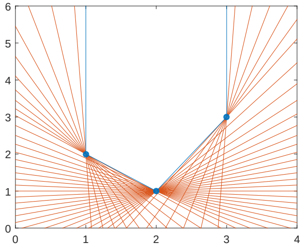


Relaxation



Relaxation

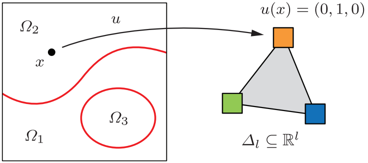
- Minimizing g^{**} we get (in this simple case) the same minimizer as when solving the original combinatorial problem! But we can now use continuous optimization methods, for example using conic programming.



- The biconjugate approach allows to systematically compute the "best" relaxation.
- Not all energies can be exactly ("tightly") relaxed like this!

Relaxation – Multi-Class Labeling

- It is possible to relax $u : \Omega \to \{1, \ldots, L\}$ to $u : \Omega \to \mathbb{R}$, but there is a better way:
- Multi-class relaxation: [Lie et al. 06, Zach et al. 08, Lellmann et al. 09, Pock et al. 09]



- Embed labels into $\mathbb{R}^L$ as $\mathcal{E} := \{e^1, \ldots, e^L\}$, relax integrality constraint to the unit simplex:

$$\Delta_L := \{x \in \mathbb{R}^L | x \geqslant 0, \sum_i x_i = 1\} = \operatorname{conv} \mathcal{E},$$

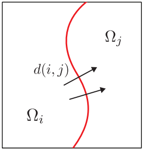
$$\min_{u:\Omega \to \Delta_L} f(u), \quad f(u) := \int_\Omega \langle u(x), s(x) \rangle dx + \int_\Omega \Psi(Du)$$

with $s(x)_i = s(i, x)$.

- The data term becomes linear!

Model – Envelope Relaxation

- We want to implement a length-based regularization: $J(\ell)$ penalizes the boundary length multiplied by an interaction potential $d(i,j)$:



- How to extend J to all functions $u : \Omega \to \Delta_L$?
- We know what the value of $J(\ell)$ should be whenever u corresponds to some ℓ, i.e., $u(x) = e^{\ell(x)}$. It is possible to show that

$$\int_\Omega \Psi(Du) = \int_{\mathcal{J}_u} \Psi((e^i - e^j)\nu^\top),$$

where $\mathcal{J}_u$ is the jump set, i and j are the labels on both sides of the jump, and ν is the normal of the boundary. We can set $\Psi((e^i - e^j)\nu^\top) = d(i,j)$ to get the length-based regularization, but it is only defined if all gradients Du of u have this particular form.

Relaxation

- We construct a regularizer of the form

$$J(u) = \int_\Omega \Psi(Du).$$

  The requirements are:
  - $\Psi((e^i - e^j)\nu^\top) = \|\nu\| d(i,j)$
  - Ψ (and therefore J) should be convex (and lower semi-continuous)
  - Ψ should not introduce additional minimizers if possible
- Use the biconjugate!

$$J(u) = \int_\Omega \Psi^{**}(Du),$$

  where

$$\Psi(M) = \begin{cases} \|\nu\| d(i,j), & \text{if } M = (e^i - e^j)\nu^\top, \\ +\infty, & \text{otherwise.} \end{cases}$$

Model – Envelope Relaxation

- Following this through, we get the following:
- $J(u)$ implicitly defined as local envelope for given d

[ChambolleCremersPock08,LellmannSchnoerr10]

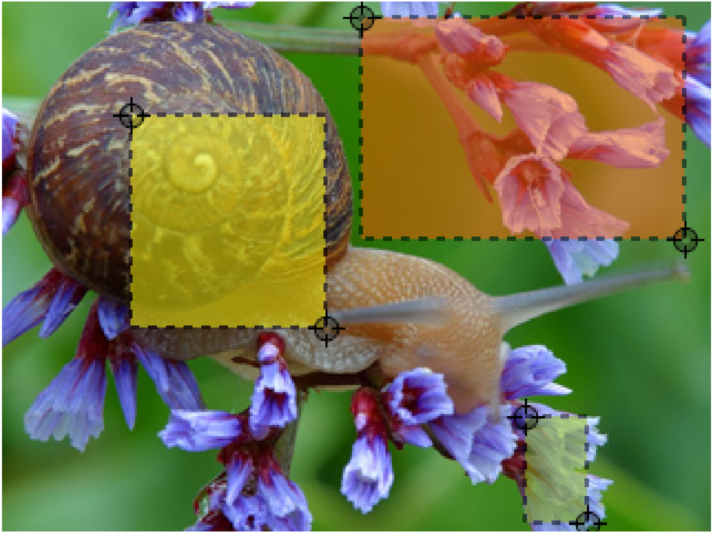
$$\begin{aligned}
J(u) &:= \sup_{v \in \mathcal{D}} \int_\Omega \langle u, \operatorname{Div} v \rangle = \int_\Omega \underbrace{\sigma_{\mathcal{D}_{\text{loc}}}(Du)}_{\Psi(Du)}, \\
\mathcal{D} &:= \left\{ v \in (C_c^\infty)^{d \times L} | v(x) \in \mathcal{D}_{\text{loc}} \; \forall x \in \Omega \right\}, \\
\mathcal{D}_{\text{loc}} &:= \left\{ (v^1, \ldots, v^L) \in \mathbb{R}^{d \times L} | \|v^i - v^j\|_2 \leqslant d(i,j) \; \forall i,j \right\}.
\end{aligned}$$

- It is also possible to use simpler but easier relaxations, e.g.,

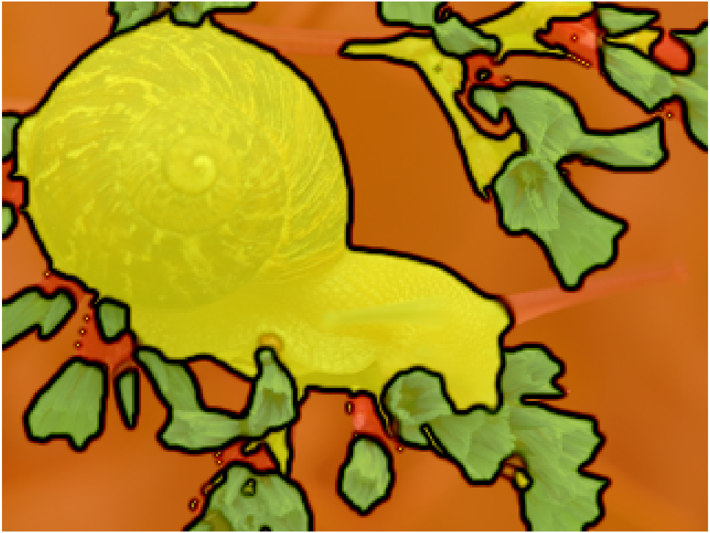
$$J(u) = \int_\Omega \|Du\|_F,$$

with the Frobenius norm $\|A\|_F = \left(\sum_{i,j} A_{ij}^2 \right)^{1/2}$.

Histogram-based segmentation



Histogram-based segmentation



Histogram-based segmentation

Numerical solution – CVX

Variational image restoration and segmentation

Jan Lellmann

Cambridge Image Analysis
Department for Applied Mathematics and Theoretical Physics
Cambridge University

Granada, May 2015

Model – Rounding

- *Fractional* solutions may occur:

- This is the cost that we pay for solving the easier relaxed problem
- Will generally happen if there is more than one solution

Rounding – Generalized Coarea Formula

- Two-class case: Generalized *coarea formula* [Strang83, ChanEsedogluNikolova06, Zach et al. 09, Olsson et al. 09]

$$f(u) = \int_0^1 f(\bar{u}_\gamma) d\gamma\,, \quad \bar{u}_\gamma := \begin{cases} e^1, & u_1(x) > \gamma, \\ e^2, & u_1(x) \leqslant \gamma. \end{cases}$$

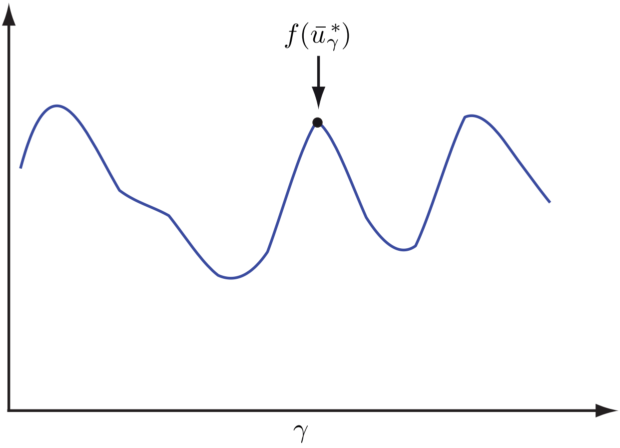
- Also: *Choquet integral, Lovász extension, levelable function*,...
- Consequence: $C = 1$, global *integral* minimizer for a.e. γ! Why? If not, then

$$\int_0^1 f(\bar{u}_\gamma^*) d\gamma > \int_0^1 f(u_{\mathcal{E}}^*) d\gamma = f(u_{\mathcal{E}}^*) \geqslant f(u^*) = \int_0^1 f(\bar{u}_\gamma^*) d\gamma,$$

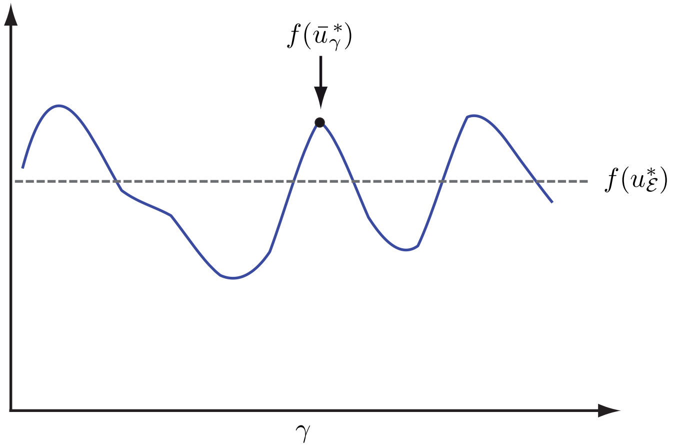
which is a contradiction to the coarea formula.

- Multi-class generalizations are possible, but we only get suboptimal solutions.

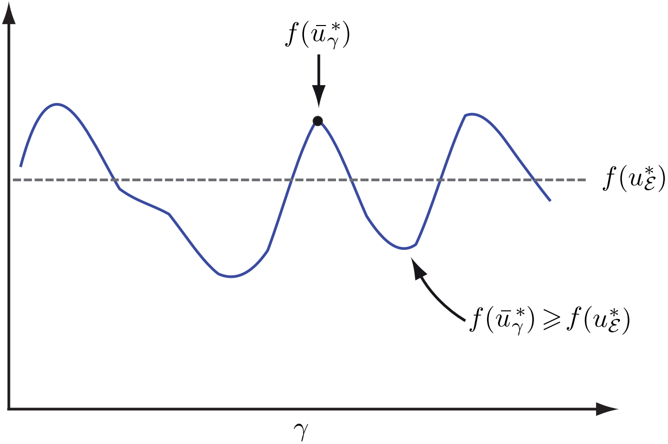
Rounding – Generalized Coarea Formula



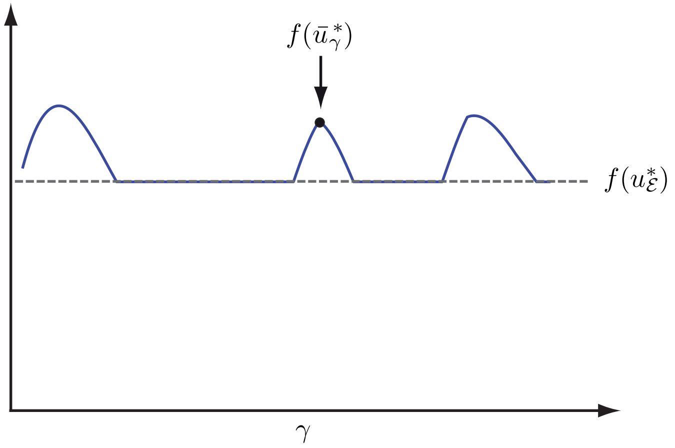
Rounding – Generalized Coarea Formula



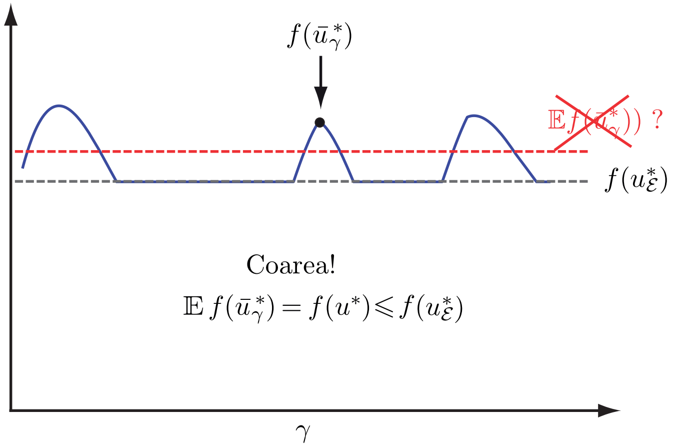
Rounding – Generalized Coarea Formula



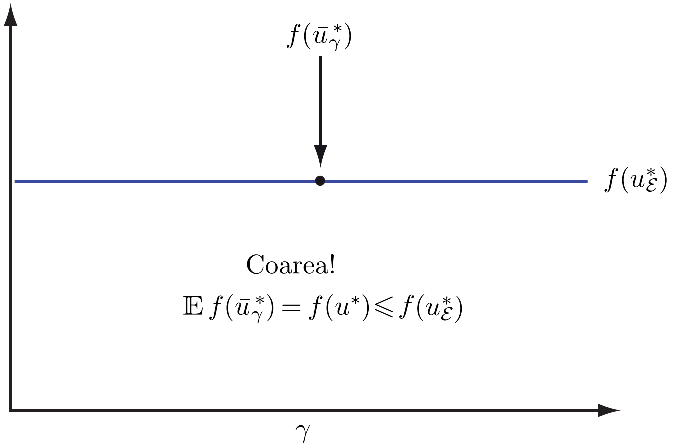
Rounding – Generalized Coarea Formula



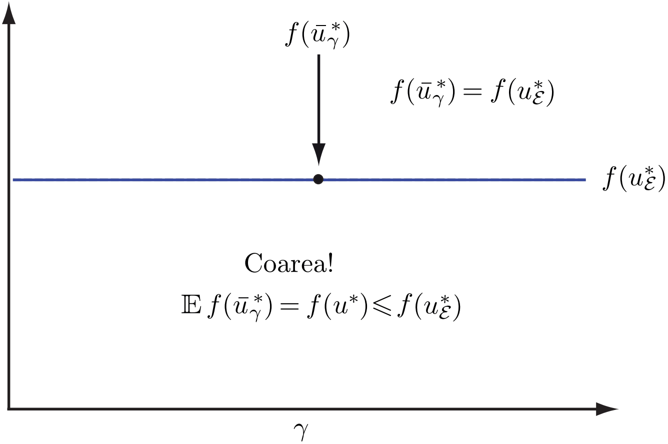
Rounding – Generalized Coarea Formula



Rounding – Generalized Coarea Formula



Rounding – Generalized Coarea Formula



Variational image restoration and segmentation

Jan Lellmann

Cambridge Image Analysis
Department for Applied Mathematics and Theoretical Physics
Cambridge University

Granada, May 2015