Inverse Problems
Variational regularisation

Clarice Poon
University of Bath

January 1, 2020

Outline

Variational regularisation

- Background
- Variational regularisation

Dual perspective

- Background II

Total variation regularisation

Variational regularisation

Let's return to Tikhonov regularisaton: The regularised solution is u_α:

$$(A^* A + \alpha \mathrm{Id}) u_\alpha = A^* f_\delta \tag{1.1}$$

One can check (do this!) that this is the first order optimality condition of

$$\min_{u \in \mathcal{U}} \|Au - f_\delta\|^2 + \frac{\alpha}{2} \|u\|^2 . \tag{1.2}$$

Since this is a convex optimisaton problem, (1.1) is a necessary and sufficient condition for the minimum of the functional (1.2).

- $\|Au - f\|^2$ is called the data fidelity term.
- $\mathcal{J}(u) \stackrel{\text{def.}}{=} \frac{1}{2} \|u\|^2$ is called the regularisation term, and penalises some unwanted features of the solution (in this case, large norm).
- α is the regularisation parameter.

Variational regularisation

We will now study more general variational regularisers of the form

$$R_\alpha f \in \operatorname{argmin}_{u \in \mathcal{U}} \mathcal{F}(Au, f_\delta) + \alpha \mathcal{J}(u). \tag{1.3}$$

where

- $A : \mathcal{U} \to \mathcal{V}$ is a bounded linear operator between Banach spaces $\mathcal{U}$ and $\mathcal{V}$
- $\mathcal{J} : \mathcal{U} \to [0, \infty)$.
- $f_\delta \in \mathcal{V}$ satisfies $\|Au - f_\delta\| \leqslant \delta$.

Banach spaces are complete, normal vector spaces.

Dual spaces

For every Banach space $\mathcal{U}$, its dual space $\mathcal{U}^*$ is the space of continuous linear functionals on $\mathcal{U}$, that is, $\mathcal{U}^* = \mathcal{L}(\mathcal{U}, \mathbb{R})$. Given $u \in \mathcal{U}$ and $p \in \mathcal{U}^*$, we write the dual product $\langle p,\, u\rangle \stackrel{\text{def.}}{=} p(u)$. The dual space is a Banach space equipped with the norm

$$\|p\|_{\mathcal{U}^*} = \sup_{u \in \mathcal{U}, \|u\|_{\mathcal{U}} \leqslant 1} \langle p,\, u\rangle.$$

The bi-dual space of $\mathcal{U} \stackrel{\text{def.}}{=} (\mathcal{U}^*)^*$. Every $u \in \mathcal{U}$ defines a continuous linear mapping on $\mathcal{U}^*$, by

$$\langle Eu,\, p\rangle \stackrel{\text{def.}}{=} \langle p,\, u\rangle.$$

$E : \mathcal{U} \to \mathcal{U}^{**}$ is well defined and is a continuous linear isometry. If E is injective, then $\mathcal{U}$ is called reflexive. Examples of reflexive Banach spaces include Hilbert spaces, L^q, ℓ^q for $q \in (1, \infty)$. We call $\mathcal{U}$ separable if there exists a countable dense subset of $\mathcal{U}$.

For any $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$, there exists a unique operator $A^* : \mathcal{V}^* \to \mathcal{U}^*$ called the adjoint of A such that for all $u \in \mathcal{U}$ and $p \in \mathcal{V}$,

$$\langle A^* p,\, u\rangle = \langle p,\, Au\rangle.$$

Background

In infinite dimensions, bounded sequences do not have to have convergent subsequences.

E.g. In ℓ^2, consider e_j the canonical basis. Then, $\|e_j\| = 1$ for all j but there does not exists $u \in \ell^2$ such that $\|e_j - u\| \to 0$.

Weak and weak-* convergence

We say that $\{u_k\} \subset \mathcal{U}$ converges weakly to $u \in \mathcal{U}$ if and only if for all $p \in \mathcal{U}^*$, we have $\langle p, u_k \rangle \to \langle p, u \rangle$.

For $\{p_k\} \subset \mathcal{U}^*$, we say $\{p_k\}$ converges weak-* to $p \in \mathcal{U}^*$ if for all $u \in \mathcal{U}$, we have $\langle p^k, u \rangle \to \langle p, u \rangle$ for all $u \in \mathcal{U}$.

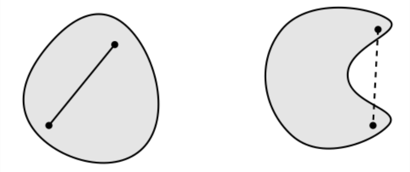
- Banach-Alaogu Theorem: Let $\mathcal{U}$ be a separable normed vector space. Then every bounded sequence has a weak-* convergent subsequence.
- Every bounded sequence in a reflexive Banach space has a weakly convergent subsequence.

We consider functionals $E : \mathcal{U} \to \bar{\mathbb{R}} \overset{\text{def.}}{=} \mathbb{R} \cup \{-\infty, +\infty\}$.

- Useful to model constraints. E.g. if $E : [-1, \infty) \to \mathbb{R}^2$ maps $x \mapsto x^2$, consider instead $\bar{E} : \mathbb{R} \to \bar{\mathbb{R}}$ defined by $\bar{E}(x) = E(x)$ for $x \in [-1, \infty)$ and $\bar{E}(x) = +\infty$ otherwise. No need to worry if $E(x + y)$ is well-defined.
- We then consider unconstrained minimisation (although the function may no longer be differentiable).
- The indicator function on a set $C \subset \mathcal{U}$ is $\iota_C \overset{\text{def.}}{=} \begin{cases} 1 & x \in C \\ +\infty & x \notin C \end{cases}$. So, we can write $\min_{u \in C} E(u) = \min_{u \in \mathcal{U}} E(u) + \iota_C(u)$.

We denote $\mathrm{dom}(E) \overset{\text{def.}}{=} \{u \in \mathcal{U} \; ; \; E(u) < \infty\}$. We say E is proper if $\mathrm{dom}(E) \neq \emptyset$.

A subset $C \subseteq \mathcal{U}$ is called convex if $\lambda u + (1-\lambda)v \in \mathcal{C}$ for all $\lambda \in (0,1)$ and $u, v \in \mathcal{C}$



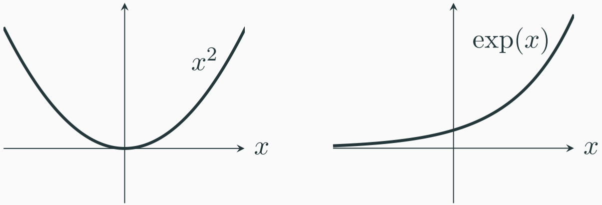
A functional $E : \mathcal{U} \to \bar{\mathbb{R}}$ is called convex if

$$E(\lambda u + (1-\lambda)v) \leqslant \lambda E(u) + (1-\lambda)E(v), \forall \lambda \in (0,1) \quad \text{and} \quad \forall u, v \in \mathrm{dom}(E), u \neq v.$$

It is called strictly convex if the inequality is strict.

Let $E : \mathcal{U} \to \bar{\mathbb{R}}$. We say that $u^* \in \mathcal{U}$ solves the minimisation problem $\min_{u \in \mathcal{U}} E(u)$ if and only if $E(u^*) \leqslant E(u)$ for all $u \in \mathcal{U}$.

A functional is called coercive if for all $u_j \in \mathcal{U}$ with $\|u_j\| \to +\infty$, we have $E(u_j) \to +\infty$. Equivalently, if $\{E(u_j)\}_j$ is bounded, then $\{u_j\}_j$ must be bounded.



Coercivity is sufficient to ensure boundedness of minimising sequences:

Lemma 1.1

Let $E : \mathcal{U} \to \bar{\mathbb{R}}$ be a proper coercive functional, bounded from below. Then, $\inf_{u \in \mathcal{U}} E(u)$ exists in $\mathbb{R}$ and there exists a minimising sequence $\{u_j\}$ such that $E(u_j) \to \inf_u E(u)$ and all minimising sequences are bounded.

Minimising functionals

Theorem 1.2 (The Direct method of Calculus)

Let $\mathcal{U}$ be a Banach space and $\tau_{\mathcal{U}}$ a topology (not necessarily the norm topology) on $\mathcal{U}$ such that bounded sequences have $\tau_{\mathcal{U}}$ convergent subsequences. Let $E : \mathcal{U} \to \bar{\mathbb{R}}$ be proper coercive and $\tau_{\mathcal{U}}$-l.s.c, and bounded from below. Then E has a minimiser.

Proof.

- The assumptions imply that there exists a bounded minimising sequence $\{u_j\}_j$.
- By assumption on the topology $\tau_{\mathcal{U}}$, there exists a subsequence u_{k_j} and $u_* \in \mathcal{U}$ which converges $\tau_{\mathcal{U}}$ to u_*.
- Due to $\tau_{\mathcal{U}}$-lsc, we have $E(u^*) \leqslant \liminf_{k\to\infty} E(u_{j_k}) = \inf_u E(u) > \infty$. Therefore, u_* is a minimiser.

□

- Key ingredient: bounded sequences have convergent subsequences.
- If $\mathcal{U}$ is a reflexive Banach space and E is a proper, bounded from below, coercive, lsc wrt weak topology, then a minimiser exists, since reflexive Banach spaces are weakly compact.
- A convex function is lsc wrt weak topology if and only if it is lsc with respect to strong topology.
- If E has at least one minimiser and is strictly convex, then the minimiser is unique: let u, v be two minimisers of E. If $u \neq v$, then

$$E(u) \leqslant E(\frac{1}{2}u + \frac{1}{2}v) < \frac{1}{2}E(u) + \frac{1}{2}E(v) \leqslant E(u)$$

which is a contradiction.

We now study the properties of

$$R_\alpha f \in \operatorname{argmin}_{u\in\mathcal{U}} \frac{1}{2}\|Au - f_\delta\| + \alpha\mathcal{J}(u).$$

as a convergent regularisation for

$$Au = f \tag{1.4}$$

where $A : \mathcal{U} \to \mathcal{V}$ is a bounded linear operator and $\mathcal{U}$, $\mathcal{V}$ are Banach spaces.

- When do minimisers exist? (i.e. well-posedness of the regularised problem)
- Are there parameter choice rules that guarantee the convergence of the minimisers to an appropriated generalised solution? (Need equivalent notions of minimal-norm solution and least squares solution)

Definition 1 ($\mathcal{J}$-minimising solutions)

Suppose that the fidelity term is such that the optimisation problem

$$\min_{u \in \mathcal{U}} \|Au - f\|$$

has a solution for any $f \in \mathcal{V}$. Let

- $u^\dagger_{\mathcal{J}} \in \operatorname{argmin}_{u \in \mathcal{U}} \mathcal{F}(Au, f)$ and
- $J(u^\dagger_{\mathcal{J}}) \leqslant \mathcal{J}(\tilde{u})$ for all $\tilde{u} \in \operatorname{argmin}_{u \in \mathcal{U}} \mathcal{F}(Au, f)$.

Then, $u^\dagger_{\mathcal{J}}$ is called a $\mathcal{J}$-minimising solution of (1.4).

Remarks

- In the following, we assume that (1.4) is solvable, i.e. for any f, there exists $u^\dagger$ such that $Au^\dagger = f$ and $\mathcal{J}(u^\dagger) < \infty$. Then, under a natural assumption that $\mathcal{F}(f,g) \geqslant 0$ and $\mathcal{F}(f,f) = 0$, it follows that $\min_{u \in \mathcal{U}} \mathcal{F}(Au, f)$ is solvable with optimum value 0.
- But, even in this case the existence of a $\mathcal{J}$-minimising solution is not guaranteed. Furthermore, even when there is existence, in general, there is no uniqueness.

Theorem 2

Let $\mathcal{U}$ and $\mathcal{V}$ be Banach spaces with topologies $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ respectively. Let $\|\cdot\|_{\mathcal{V}}$ be $\tau_{\mathcal{V}}$-lsc. Suppose that $Au = f$ has a solution with finite $\mathcal{J}$-value. Assume that

(i) $A : \mathcal{U} \to \mathcal{V}$ *is* $\tau_{\mathcal{U}} \to \tau_{\mathcal{V}}$ *continuous.*

(ii) $\mathcal{J} : \mathcal{U} \to (0, +\infty]$ *is proper,* $\tau_{\mathcal{U}}$*-lsc and its non-empty sublevel-sets* $\{u \in \mathcal{U} \ ; \ \mathcal{J}(u) \leqslant C\}$ *are* $\tau_{\mathcal{U}}$*-sequentially compact*

Then,

(i') *there exists a* $\mathcal{J}$*-minimising solution* $u^{\dagger}_{\mathcal{J}}$.

(ii') *for any fixed* $\alpha > 0$ *and* $f_\delta \in \mathcal{V}$, *there exists a minimiser of* $u^{\alpha}_{\delta} \in \operatorname{argmin}_u \|Au - f_\delta\|^2 + \alpha \mathcal{J}(u)$.

Theorem 2

Let $\mathcal{U}$ and $\mathcal{V}$ be Banach spaces with topologies $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ respectively. Let $\|\cdot\|_{\mathcal{V}}$ be $\tau_{\mathcal{V}}$-lsc. Suppose that $Au = f$ has a solution with finite $\mathcal{J}$-value. Assume that

(i) *$A : \mathcal{U} \to \mathcal{V}$ is $\tau_{\mathcal{U}} \to \tau_{\mathcal{V}}$ continuous.*

(ii) *$\mathcal{J} : \mathcal{U} \to (0, +\infty]$ is proper, $\tau_{\mathcal{U}}$-lsc and its non-empty sublevel-sets $\{u \in \mathcal{U} \; ; \; \mathcal{J}(u) \leqslant C\}$ are $\tau_{\mathcal{U}}$-sequentially compact*

Then,

(i') *there exists a $\mathcal{J}$-minimising solution $u^{\dagger}_{\mathcal{J}}$.*

(ii') *for any fixed $\alpha > 0$ and $f_\delta \in \mathcal{V}$, there exists a minimiser of $u^{\alpha}_{\delta} \in \operatorname{argmin}_u \|Au - f_\delta\|^2 + \alpha \mathcal{J}(u)$.*

For (i'). Let $\mathbb{L} \stackrel{\text{def.}}{=} \{u \; ; \; Au = f\}$. Consider $\inf_{u \in \mathbb{L}} \mathcal{J}(u)$.

- $\mathbb{L}$ is nonempty by assumption and closed by continuity of A.
- Since $\mathcal{J} \geqslant 0$, there exists a minimising sequence u_n. By compactness of sublevel sets, there exists a subsequence u_{n_k} which $\tau_{\mathcal{U}}$ converges to u_*.

Theorem 2

Let $\mathcal{U}$ and $\mathcal{V}$ be Banach spaces with topologies $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ respectively. Let $\|\cdot\|_{\mathcal{V}}$ be $\tau_{\mathcal{V}}$-lsc. Suppose that $Au = f$ has a solution with finite $\mathcal{J}$-value. Assume that

(i) *$A : \mathcal{U} \to \mathcal{V}$ is $\tau_{\mathcal{U}} \to \tau_{\mathcal{V}}$ continuous.*

(ii) *$\mathcal{J} : \mathcal{U} \to (0, +\infty]$ is proper, $\tau_{\mathcal{U}}$-lsc and its non-empty sublevel-sets $\{u \in \mathcal{U} \; ; \; \mathcal{J}(u) \leqslant C\}$ are $\tau_{\mathcal{U}}$-sequentially compact*

Then,

(i') *there exists a $\mathcal{J}$-minimising solution $u^{\dagger}_{\mathcal{J}}$.*

(ii') *for any fixed $\alpha > 0$ and $f_\delta \in \mathcal{V}$, there exists a minimiser of $u^{\alpha}_{\delta} \in \operatorname{argmin}_u \|Au - f_\delta\|^2 + \alpha \mathcal{J}(u)$.*

For (i'). Let $\mathbb{L} \overset{\text{def.}}{=} \{u \; ; \; Au = f\}$. Consider $\inf_{u \in \mathbb{L}} \mathcal{J}(u)$.

- $\mathbb{L}$ is nonempty by assumption and closed by continuity of A.
- Since $\mathcal{J} \geqslant 0$, there exists a minimising sequence u_n. By compactness of sublevel sets, there exists a subsequence u_{n_k} which $\tau_{\mathcal{U}}$ converges to u_*.

Theorem on convergent regularisation

Theorem 3

Under the assumptions of Theorem 2, if $\alpha = \alpha(\delta)$ is such that $\delta^2/\alpha(\delta) \to 0$ as $\delta \to 0$, then $u_\delta \stackrel{\text{def.}}{=} u_\delta^{\alpha(\delta)}$ converges (up to a subsequence) $\tau_\mathcal{U}$ to $u^\dagger_\mathcal{J}$ as $\mathcal{J}$ minimising solution and $\mathcal{J}(u_\delta) \to \mathcal{J}(u^\dagger_\mathcal{J})$.

- Since u_δ is a minimiser:
  - $\|Au_\delta - f_\delta\|^2 + \alpha(\delta)\mathcal{J}(u_\delta) \leqslant \frac{1}{2}\left\|Au^\dagger_\mathcal{J} - f_\delta\right\|^2 + \alpha(\delta)\mathcal{J}(u^\dagger_\mathcal{J})$
  - $J(u_\delta) \leqslant \mathcal{J}(u^\dagger_\mathcal{J}) + \frac{\delta^2}{2\alpha(\delta)}$.
- by compactness of the sublevel sets of $\mathcal{J}$, up to a subsequence u_{δ_n} converges to u_* as $\delta_n \to 0$. By continuity of A, $Au_{\delta_n} \stackrel{\tau_\mathcal{V}}{\to} Au_*$.
- $Au_* = f$ follows by lsc of $\|\|_\mathcal{V}$ wrt $\tau_\mathcal{V}$ and by minimality of u_{δ_n}:

$$\begin{aligned}\frac{1}{2}\|Au_* - f\|^2 &\leqslant \liminf \|Au_{\delta_n} - f_\delta\|^2 \leqslant \liminf \frac{1}{2}\|Au_{\delta_n} - f_\delta\| + \alpha(\delta_n)\mathcal{J}(u_{\delta_n}) \\ &\leqslant \liminf \frac{1}{2}\left\|Au^\dagger_\mathcal{J} - f_\delta\right\| + \alpha(\delta_n)\mathcal{J}(u^\dagger_\mathcal{J}) = 0\end{aligned}$$

- Finally
  $\mathcal{J}(u_*) \leqslant \liminf_{n\to\infty} \mathcal{J}(u_{\delta_n}) \leqslant \liminf_{n\to\infty} \mathcal{J}(u^\dagger_\mathcal{J}) + \frac{\delta_n^2}{2\alpha(\delta_n)} = \mathcal{J}(u^\dagger_\mathcal{J})$.

Given $g \in \mathbb{R}^N$, there are two components to (linear) inverse problems:

1. A data model: $g = Tu_0 + n$ where $u_0 \in \mathbb{R}^N$ is the underlying object to be recovered, T is some linear transform (e.g. a blurring operator, a subsampled Fourier transform, or the identity matrix), and n is the noise. Typically, the entries in n are assumed to be Gaussian distributed with mean 0 and variance σ^2.
2. An a-priori probability density: $P(u) = e^{-p(u)}$. This represents the idea that we have of the solution.

Bayesian viewpoint of variational methods

By Bayes' rule, the posteriori probability of u knowing g is

$$P(u|g)P(g) = P(g|u)P(u),$$

where $P(g|u) = \exp\left(-\frac{1}{\sigma^2}\|g - Tu\|_2^2\right)$. So,

$$P(u|g) = \frac{\exp\left(-\frac{1}{\sigma^2}\|g - Tu\|_2^2 - p(u)\right)}{P(g)},$$

Bayesian viewpoint of variational methods

By Bayes' rule, the posteriori probability of u knowing g is

$$P(u|g)P(g) = P(g|u)P(u),$$

where $P(g|u) = \exp\left(-\frac{1}{\sigma^2}\|g - Tu\|_2^2\right)$. So,

$$P(u|g) = \frac{\exp\left(-\frac{1}{\sigma^2}\|g - Tu\|_2^2 - p(u)\right)}{P(g)},$$

The maximum a posteriori (MAP) reconstruction is:

$$u^* \in \underset{u}{\operatorname{argmax}}\, P(u|g).$$

Equivalently,

$$u^* \in \underset{u}{\operatorname{argmin}}\, p(u) + \frac{1}{\sigma^2}\|g - Tu\|_2^2.$$

Other choices of noise distributions:

- Additive Laplace noise $e^{-\frac{1}{\sigma^2}\|g - Tu\|_1}$ with corresponding data fidelity term $\|Tu - g\|_1$
- Poisson noise $\prod_{i,j} \frac{u_{i,j}^{g_{i,j}}}{g_{i,j}!} e^{-u_{i,j}}$ with data fidelity term $\int u - g\log(u)$.

Let $\mathcal{U}$ be a Hilbert space and $\mathcal{J}(u) = \|u\|^2$.

- this is weakly-lsc and bounded sequences have weakly convergent subsequences. Note that in (iii'), since Hilbert spaces satisfy the Radon Riesz property, we have strong convergence as well as weak convergence of solutions.

Classical examples are Sobolev spaces such as $H^1 = W^{1,2}$ and L^2. One can show that in 1D, H^1 consists only of continuous functions and therefore, regularised solutions must be continuous. So, $\mathcal{J}(u) = \|u\|_{H^1}$ is sometimes referred to as the smoothing functional.

Examples of regularisers

Let $\mathcal{U} = \ell^2$ be the space of square summable sequences. Let $\mathcal{J}(u) = \|u\|_1 = \sum_j |u_j|$.

- $\mathcal{J}$ is weakly lsc in ℓ^2.
- We have $\|\cdot\|_2 \leqslant \|\cdot\|_1$, so $\mathcal{J}(u) \leqslant C$ implies $\|u\|_2 \leqslant C$ and bounded sequences have weakly convergent subsequences in ℓ^2. So, the sublevel-sets of $\mathcal{J}$ are weakly sequentially compact in ℓ^2.

One popular regularisation is the *lasso*:

$$\min_u \frac{1}{2} \|Au - f\|_2^2 + \alpha \|u\|_1 .$$

Outline

Towards convergence rates

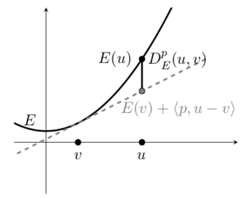
We have established convergence of a regularised solution u_δ to a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ as $\delta \to 0$. We now establish results on the *speed* of convergence.

Convergence rates are typically derived via analysis of the *dual problem* and studied in terms of the *Bregman distances* associated with the (convex) regularisation functional $\mathcal{J}$.

Definition 4

Given a convex functional $\mathcal{J}$, $u, v \in \mathcal{U}$ such that $\mathcal{J}(v) < \infty$ and $q \in \partial\mathcal{J}(v)$, the generalised Bregman distance is given by

$$\mathcal{D}^q_{\mathcal{J}}(u, v) = \mathcal{J}(u) - \mathcal{J}(v) - \langle q,\, u - v\rangle. \tag{2.1}$$



Example: For $\mathcal{J}(u) = \frac{1}{2}\|u\|^2$, the subgradient at $v$is $q = v$, so

$$\mathcal{D}^v_{\mathcal{J}}(u, v) = \frac{1}{2}\|u\|^2 - \frac{1}{2}\|v\|^2 - \langle v,\, u - v\rangle = \frac{1}{2}\|u\|^2 + \frac{1}{2}\|v\|^2 - \langle v,\, u\rangle = \frac{1}{2}\|u - v\|^2 .$$

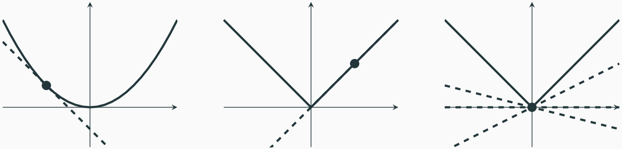
For convex functionals, we can generalise the concept of a derivative for non-differentiable functions.

Definition 5

A functional $E : \mathcal{U} \to \bar{\mathbb{R}}$ is called subdifferentiable at $u \in \mathcal{U}$ if there exists an element $p \in \mathcal{U}^*$ such that $E(v) \geqslant E(u) + \langle p, v - u\rangle$ for all $v \in \mathcal{U}$. We call p a subgradient at u. The collection of all subgradients at u

$$\partial E(u) \stackrel{\text{def.}}{=} \{p \in \mathcal{U}^* \; ; \; E(v) \geqslant E(u) + \langle p, v - u\rangle, \forall v \in \mathcal{U}\}$$

is called the subdifferential of E at u.



Let $E : \mathbb{R} \to \mathbb{R}$ be $E(u) = |u|$. Then, $\partial E(u) = \begin{cases} \text{sign}(u) & u \neq 0 \\ [-1, 1] & u = 0 \end{cases}$

- If E is differentiable at u, then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$.
- Let $E : \mathcal{U} \to \bar{\mathbb{R}}$ and $F : \mathcal{U} \to \bar{\mathbb{R}}$ be proper lsc convex functions and suppose that there exists $u \in \mathrm{dom}(E) \cap \mathrm{dom}(F)$ such that E is continuous at u. Then $\partial(E + F) = \partial E + \partial F$.
- Let E be convex. Then, u is a minimises E if and only if $0 \in \partial E(u)$.
- If $E : \mathcal{U} \to \bar{\mathbb{R}}$ is a proper convex function and $u \in \mathrm{dom}(E)$, then $\partial E(u)$ is a weak-* compact convex subset of $\mathcal{U}^*$.

The convex conjugate

Let V be a real topological vector space and let V^* be its dual.

Definition 6

Given $F : V \to (-\infty, +\infty]$, its convex conjugate is $F^* : V^* \to (-\infty, +\infty]$ defined by

$$F^*(y) \overset{\text{def.}}{=} \sup_{x \in V} \{\langle x, y\rangle - F(x)\}.$$

- F^* is convex regardless of whether F is convex.
- We have the Fenchel Young inequality: $\langle x, y\rangle \leqslant F(x) + F^*(y)$,
- if F is convex and lower semi-continuous, then $F^{**} = F$.
- if F is convex, then $y \in \partial F(x)$ if and only if $F(x) + F^*(y) = \langle x, y\rangle$.

The convex conjugate – Examples

(a) if $F(x) = \frac{1}{2}\|x\|^2$ and V is a Hilbert space, then $F^*(y) = \frac{1}{2}\|y\|^2$:

- $F^*(y) = \sup_x \langle x, y\rangle - \frac{1}{2}\|x\|^2 \leqslant \frac{1}{2}\|y\|^2$.
- Setting $x \stackrel{\text{def.}}{=} y$ in the supremum above yields $F^*(y) \geqslant \frac{1}{2}\|y\|^2$.

(b) If $F(x) = \|x\|$ and $\|\cdot\|_*$ is its dual norm, then

$$F^*(y) = \begin{cases} 0 & \|y\|_* \leqslant 1 \\ +\infty & \text{otherwise.} \end{cases}$$

(c) If $F = \iota_K$ (takes value 0 for $x \in K$ and $+\infty$ otherwise) with K being a convex set, then $F^*(y) = \sup_{x \in K} \langle x, y\rangle$.

Absolutely one-homogeneous functionals

A functional $E : \mathcal{U} \to \bar{\mathbb{R}}$ is absolutely one-homogeneous if $E(\lambda u) = |\lambda|\, E(u)$ for all $\lambda \in \mathbb{R}$ and $u \in \mathcal{U}$. Clearly $E(0) = 0$.

One example is the total variation functional.

- Let E be convex, absolutely one-homogeneous and let $p \in \partial E(u)$. Then $E(u) = \langle p,\, u\rangle$.
- Let E be proper, convex, lsc, absolutely one-homogeneous. Then, E^* is the characteristic function of the convex set $\partial E(0)$.
- for any $u \in \mathcal{U}$, $p \in \partial E(u)$ if and only if $p \in \partial E(0)$ and $E(u) = \langle p,\, u\rangle$.

Primal and dual formulations

Let V, Y be real topological vector spaces with duals V^* and Y^*. Let $y \in Y$ and $b_j \in \mathbb{R}$ for $j = 1, \ldots, M$. Consider **the primal problem**:

$$\min_{x \in V} F_0(x) \text{ subject to } Ax = y, \tag{2.2}$$

$$F_j(x) \leqslant b_j,\ j \in [M], \tag{2.3}$$

where $F_0 : V \to (-\infty, +\infty]$ is called the objective function and $F_j : V \to (-\infty, +\infty]$ for $j \in [M]$ are called the constraint functions. $A : V \to Y$ is a continuous linear functional. The set $K \overset{\text{def.}}{=} \{x \in V\ ;\ Ax = y, F_j(x) \leqslant b_j\}$ is called the admissible set.

The **Lagrange function** is defined for $x \in V$, $\xi \in Y^*$ and $\nu \in \mathbb{R}^M$ with $\nu_\ell \geqslant 0$ for all $\ell \in [M]$ by

$$L(x, \xi, \nu) \overset{\text{def.}}{=} F_0(x) + \langle \xi, Ax - y \rangle + \sum_{\ell=1}^{M} \nu \left(F_\ell(x) - b_\ell \right).$$

The variables ξ and ν are called the **Lagrange multipliers**.

The Lagrange dual function is defined as

$$H(\xi,\nu) \stackrel{\text{def.}}{=} \inf_{x\in V} L(x,\xi,\nu), \qquad \xi \in Y^*,\ \nu \in \mathbb{R}^M_{\geqslant 0}.$$

If $x \mapsto L(x,\xi,\nu)$ is unbounded from below, then we write $H(\xi,\nu) = -\infty$.

- The dual function is always concave since it is the pointwise infimum of a family of affine functions.
- We have $H(\xi,\nu) \leqslant \inf_{x\in K} F_0(x)$ for all $\xi \in Y^*$ and $\nu \in \mathbb{R}^M_{\geqslant 0}$. Indeed, we have $H(\xi,\nu) \leqslant \inf_{x\in K} L(x,\xi,\nu)$, and note that given any $x \in K$, we have $Ax - y = 0$ and $F_\ell(x) - b_\ell \leqslant 0$, so $L(x,\xi,\nu) \leqslant F_0(x)$.

So, $H(\xi,\nu)$ serves as a lower bound for the infimum of F_0 over K, and since we want this lower bound to be as tight as possible, it makes sense to consider

$$\sup_{\xi\in Y^*,\nu\in\mathbb{R}^M} H(\xi,\nu) \text{ subject to } \nu_\ell \geqslant 0,\ \ell \in [M]. \tag{2.4}$$

This optimisation problem is called the **dual problem** and (2.2) is called the **primal problem**.

- If D^* is the supremum of (2.4) and P^* is the infimum of (2.2), then we have in general $D^* \leqslant P^*$ (this is called **weak duality**). and $P^* - D^*$ is called the duality gap.
- When $D^* = P^*$, then we say we have **strong duality**.

Primal and dual formulations

Consider now $\inf_{x \in V} E(Ax) + F(x)$, where $E : Y \to (-\infty, +\infty]$ and $F : V \to (-\infty, +\infty]$ are convex functionals, and $A : V \to Y$ is a continuous linear operator. This is equivalent to

$$\inf_{z \in Y, x \in V} E(z) + F(x) \text{ subj. to } Ax = z$$

the Lagrange dual is for $\xi \in Y^*$ as

$$\begin{aligned} H(\xi) &= \inf_{x,z} \{E(z) + F(x) + \langle \xi, Ax - z \rangle\} \\ &= \inf_{x,z} \{E(z) + F(x) + \langle A^* \xi, x \rangle - \langle \xi, z \rangle\} \\ &= -\sup_{z \in Y} \langle \xi, z \rangle - E(z) - \sup_{x \in V} \langle -A^* \xi, x \rangle - F(x) \\ &= -E^*(\xi) - F^*(-A^* \xi). \end{aligned}$$

So, the dual problem is

$$\sup_{\xi \in Y^*} -E^*(\xi) - F^*(-A^* \xi)$$

Theorem 2.1 (Strong duality)

Suppose that E and F are proper convex functionals, there exists $u_0 \in V$ such that $F(u_0) < \infty$, $E(Au_0) < \infty$ and E is continuous at Au_0. Then, strong duality holds and there exists at least one dual optimal solution. Moreover, if p^* is a primal optimal solution and d^* is a dual optimal solution, then

$$Ap^* \in \partial E^*(d^*) \quad \textit{and} \quad A^* d^* \in -\partial F(p^*)$$

Primal and dual formulations

We are interested in the case

$$\min_u \frac{1}{2} \|Au - f_\delta\|^2 + \alpha \mathcal{J}(u)$$

So, $E(Au) = \frac{1}{2} \|Au - f_\delta\|^2$ and $F(u) = \alpha \mathcal{J}(u)$.

- $E^*(v) = -\langle v,\, f_\delta \rangle + \frac{1}{2} \|v\|^2$.
- If $\mathcal{J}$ is absolute one-homogeneous, then $\mathcal{J}^*(v) = \iota_K$ where $K = \partial J(0)$, and $(\alpha J)^*(v) = \alpha J^*(\alpha^{-1} v)$.

Therefore, the dual problem is

$$\sup_v \langle v,\, f_\delta \rangle - \frac{1}{2} \|v\|^2 + \iota_K \left(\frac{A^* v}{\alpha} \right) = \sup_{v : A^* v \in \partial \mathcal{J}(0)} \alpha \left(\langle v,\, f_\delta \rangle - \frac{\alpha}{2} \|\alpha v\|^2 \right). \quad (2.5)$$

If p_δ and u_δ are dual and primal solutions, then the optimality conditions take the form

$$A^* p_\delta \in \partial \mathcal{J}(u_\delta) \quad \text{and} \quad p = \frac{f_\delta - Au_\delta}{\alpha}$$

NB: the dual solution is unique since it is the projection onto a closed convex set.

The limit primal and dual problems

Formal limits problems as $\delta \to 0$ are

$$\inf_{u: Au = f} \mathcal{J}(u) = \inf_{u \in \mathcal{U}} \iota_{\{f\}}(Au) + \mathcal{J}(u) \tag{2.6}$$

and

$$\sup_{p: A^* p \in \partial \mathcal{J}(0)} \langle f, p \rangle = -\inf_{p} \langle -f, p \rangle + \iota_{\partial J(0)}(A^* p) \tag{2.7}$$

Lemma 7

For $J : \mathcal{U} \to [0, \infty]$ absolute one-homogeneous and coercive, we have $0 \in \mathrm{int}(\partial J(0))$.

Proof.

Indeed, if not, then there exists e_n and u_n with $\|e_n\| \to 0$ such that $J(u_n) < \langle e_n, u_n \rangle$. Since J is one-homogeneous, we can assume that $\|u_n\| = 1$. Therefore, $\lim_{n\to\infty} J(u_n) \leqslant \lim_{n\to\infty} \|e_n\| \, \|u_n\| = 0$. Letting $\lambda_n = 1/J(u_n)$, we have $\|\lambda_n u_n\| \to +\infty$ but $J(\lambda_n u_n) = 1$. Contradiction since J is coercive.

$\square$

The source condition

- We can apply Theorem 2.1 to (2.7) with $F = \iota_{\partial J(0)}$ and $E = \langle -f, \cdot \rangle$. Cleary, $E(0) = 0$, $F(A^*0) = 0$, and F is continuous at 0. In this case, we have strong duality and (2.6) has at least one solution.
- However, unlike the case where $\alpha > 0$, there is no guarantee that a dual solution to (2.7) exists, and it may not be unique if it does exist.
- If a dual solution p exists, then it is related to any primal solution u by $A^*p \in \partial J(u)$.

Definition 8

We say that a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ satisfies the source condition if there exists $p^\dagger \in \mathcal{V}$ such that $A^*p^\dagger \in \partial\mathcal{J}(u^\dagger)$.

We will establish convergence rates under the source condition.

What is the behaviour of p_δ as $\delta \to 0$?

Necessity of the source condition

Theorem 9

Suppose that the conditions of Theorem 2 are satisfied with $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ being weak topologies in $\mathcal{U}$ and $\mathcal{V}$ respectively. Suppose that p_δ is uniformly bounded in δ. Then, there exists $p^\dagger$ such that $A^* p^\dagger \in \partial J(u^\dagger)$.

Proof.

- Consider a sequence $\delta_n \to 0$. Since p_δ is uniformly bounded in δ, there exists a weakly convergent subsequence $p_{\delta_n} \rightharpoonup p_0$. Since A^* is weak-weak continuous, we get $A^* p_{\delta_n} \rightharpoonup A^* p_0$.
- Since $A^* p_\delta \in \partial J(0)$, a weakly closed set, we have $A^* p_0 \in \partial J(0)$.
- Since J is absolute one-homogeneous and $A^* p_\delta \in \partial \mathcal{J}(u_\delta)$, we have $\langle A^* p_\delta, u_\delta \rangle = \mathcal{J}(u_\delta)$.

  $$\langle A^* p_\delta, u_\delta \rangle = \langle A^* p_\delta, u^\dagger_{\mathcal{J}} \rangle + \langle p_\delta, Au_\delta - Au^\dagger_{\mathcal{J}} \rangle = \underbrace{\langle A^* p_\delta, u^\dagger_{\mathcal{J}} \rangle}_{\to \langle A^* p_0, u^\dagger_{\mathcal{J}} \rangle} + \underbrace{\langle p_\delta, Au_\delta - f \rangle}_{\to 0}$$

  Also, $\mathcal{J}(u_\delta) \to \mathcal{J}(u^\dagger_{\mathcal{J}})$. Therefore, $\langle A^* p_0, u^\dagger_{\mathcal{J}} \rangle = \mathcal{J}(u^\dagger_{\mathcal{J}})$.
- therefore, $A^* p_0 \in \partial \mathcal{J}(u^\dagger_{\mathcal{J}})$ and p_0 is a dual solution.

□

The source condition implies dual convergence

Theorem 2.2

Suppose that the source condition holds at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$. Then, p_α the solution to (2.5) with data f is uniformly bounded in α. Moreover, $p_\alpha \to p^\dagger$ strongly in $\mathcal{V}$ as $\alpha \to 0$, where $p^\dagger$ is a solution to (2.7) with smallest norm.

Proof.

- Let p_α be a solution to to (2.5) with $f_\delta = f$, we have
$$\langle f,\, p_\alpha \rangle - \frac{\alpha}{2} \|p_\alpha\|^2 \geqslant \langle f,\, p^\dagger \rangle - \frac{\alpha}{2} \left\|p^\dagger\right\|^2, \tag{2.8}$$
and $p^\dagger$ being a solution to (2.7) implies that $\langle f,\, p^\dagger \rangle \geqslant \langle f,\, p_\delta \rangle$. So, $\left\|p^\dagger\right\| \geqslant \|p_\alpha\|$.
- We may extract a subsequence such that $p_{\alpha_{n_k}}$ weakly converges to p_* (recall that the closed unit ball of a Hilbert space is weakly sequentially compact). Taking the limit of $\lambda \to 0$ in (2.8) yields $\langle f,\, p_* \rangle \geqslant \langle y,\, p^\dagger \rangle$.
- Note that $A^* p_{\alpha_{n_k}}$ converges weakly to $A^* p_*$, and so $A^* p_* \in \partial \mathcal{J}(0)$ (since this is a weakly closed set). So, p_* is a solution to (2.7).

Proof.

- Finally, p_* is the solution of minimal norm since

$$\|p_*\| \leqslant \liminf_k \left\|p_{\alpha_{n_k}}\right\| \leqslant \left\|p^\dagger\right\|,$$

and hence, $p_* = p^\dagger$, $\left\|p_{\alpha_{n_k}}\right\| \to \left\|p^\dagger\right\|$ and $p_{\alpha_{n_k}} \to p_0$ strongly in $\mathcal{H}$. This implies $\lim_{\delta\to 0}\left\|p_\alpha - p^\dagger\right\| = 0$, since otherwise, we can extract a subsequence p_{α_k} such that $\left\|p_{\alpha_k} - p^\dagger\right\| > \varepsilon$ and by the above argument, extract a further subsequence which converges strongly to $p^\dagger$.

□

Note: since the solution to (2.5) with f_δ is $P_K(f_\delta/\alpha)$ the orthogonal projection onto $\{p\ ;\ A^* p \in \partial\mathcal{J}(0)\}$, we have

$$\|p_\alpha - p_\delta\| = \|P_K(f/\alpha) - P_K(f_\delta/\alpha)\| \leqslant \delta/\alpha \leqslant C.$$

So, $\|p_\delta\|$ is also uniformly bounded in δ and converges to $p^\dagger$ as $\delta/\alpha(\delta) \to 0$.

Theorem 2.3

Assume that the source condition is satisfied at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ and u_δ be a regularised solution. Then, $D^{p^\dagger}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) \leqslant \frac{1}{2\alpha}\left(\delta + \alpha \left\| p^\dagger \right\|\right)^2$.

Proof.

Since u_δ is a minimiser, $\alpha \mathcal{J}(u_\delta) + \frac{1}{2}\left\| Au_\delta - f_\delta \right\| \leqslant \alpha \mathcal{J}(u^\dagger_{\mathcal{J}}) + \frac{1}{2}\left\| Au^\dagger_{\mathcal{J}} - f_\delta \right\|$.

- $\alpha D^{p^\dagger}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) + \frac{1}{2}\left\| Au_\delta - f_\delta \right\| + \alpha \langle A^* p^\dagger,\, u_\delta - u^\dagger_{\mathcal{J}} \rangle \leqslant \frac{\delta^2}{2}$.
- LHS is equal to

$$\frac{1}{2}\left\| Au_\delta - f_\delta + \alpha p^\dagger \right\|^2 + \alpha D^{p^\dagger}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) - \frac{\alpha^2}{2}\left\| p^\dagger \right\|^2 + \alpha \langle p^\dagger,\, f_\delta - f_\dagger \rangle.$$

- Rearranging and by Cauchy-Schwarz:

$$D^{p^\dagger}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) \leqslant \frac{1}{2\alpha}\left(\delta^2 + \alpha^2 \left\| p^\dagger \right\|^2 + 2\alpha \left\| p^\dagger \right\| \delta\right).$$

□

Outline

Total variation regularisation

For imaging, one may consider $F(u) = \frac{1}{2}\int_\Omega |\nabla u|^2$. One can check that u is a minimizer if and only if

$$T^* Tu - T^* g - \lambda \Delta u = 0,$$

with Neumann boundary condition $\nabla u \cdot \eta = 0$ on $\partial\Omega$ where η is the outward unit normal to $\partial\Omega$.

- Intuition is to encourages solutions with small gradient which best fit the observation data g, so noise is removed.
- leads to oversmooth reconstructions as Δ has very strong isotropic smoothing properties.

Moreover, as a consequence of classical Sobolev embedding theorems, such functions cannot exhibit discontinuities across hypersurfaces. In 2D, this corresponds to no discontinuities across lines. To offer a quick (formal) justification: if $0 < s < t < 1$, $u : [0,1] \to \mathbb{R} \in W^{1,2}(0,1)$, then

$$u(t) - u(s) = \int_s^t u'(r)\mathrm{d}r \leqslant \sqrt{t-s}\sqrt{\int_s^t |u'(r)|^2 \, \mathrm{d}r} \leqslant \sqrt{t-s}\, \|u\|_{W^{1,2}}.$$

So, u is Hölder-1/2 continuous. This is especially problematic because it is the key information about an image is encoded in its edges!

Total variation regularisation

Rudin, Osher and Fatemi introduced the total variation functional for image processing:

$$F(u) = \int_\Omega |\nabla u| \,.$$

- This functional is well defined on $W^{1,1}(\Omega)$.
- For $u \in W^{1,1}([a,b])$, define continuous function $\tilde{u}(x) - \tilde{u}(a) = \int_a^x u'(t)\mathrm{d}t$ which coincides with u a.e.. So, functions in $W^{1,1}([a,b])$ cannot have discontinuities, and given $f \in W^{1,1}([a,b]^2)$, since $f(\cdot, x) \in W^{1,1}([a,b])$ for a.e. x, images cannot have jumps across vertical/horizontal boundaries.

Key point: is that F is well defined for a more general class of functions which can have discontinuities. Furthermore, as the resultant variational problem is now convex, we can apply some standard numerical solvers.

$$\min_u \mathcal{J}(u) + \|Ku - b\|^2, \quad \text{where} \quad Ku = h \star u$$

$\mathcal{J}(x) = \|Dx\|_2^2$

$\mathcal{J}(x) = \|Dx\|_1$

We shall see in this example that not only can $\int |\nabla u|$ be extended to a larger class of functions where edges are permitted, it is actually necessary to do so.

Consider

$$\min_{u \in W^{1,1}([0,1])} \mathcal{E}(u), \qquad \mathcal{E}(u) = \lambda \int_0^1 \left|u'(t)\right| \mathrm{d}t + \int_0^1 |u(t) - g(t)|^2 \, \mathrm{d}t,$$

where $g = \chi_{(1/2,1]}$.

We will show that this minimization problem does not have a solution in $W^{1,1}$.

Motivating example

Let u be a minimizer.

- **Maximum/minimum principles** $u \leqslant 1$ a.e.:
  Let $v \in \min\{u, 1\}$. Then,
  - $v' = u'$ on $\{u < 1\}$ and $v' = 0$ on $\{u \geqslant 1\}$. Therefore, $\int |v'| \leqslant \int |u'|$.
  - Since $g \leqslant 1$, $\|v - g\|^2 \leqslant \|u - g\|^2$.

  So, $\mathcal{E}(v) \leqslant \mathcal{E}(u)$ and this inequality is strict if $v \neq u$. Similarly, $u \geqslant 0$ a.e..
- **'Symmetry'** Note that $g(t) = 1 - g(1-t)$. Let $\tilde{u} = 1 - u(1-t)$. Then $\|\tilde{u} - g\|^2 = \|u - g\|^2$ and $\|\tilde{u}'\|_1 = \|u'\|_1$. So, $\mathcal{E}(\tilde{u}) = \mathcal{E}(u)$.
  Also,
  $$\mathcal{E}\left(\frac{\tilde{u} + u}{2}\right) \leqslant \frac{1}{2}\mathcal{E}(\tilde{u}) + \frac{1}{2}\mathcal{E}(u) = \mathcal{E}(u)$$
  and by strict convexity of $\|\cdot\|_2^2$, this inequality is strict if $\tilde{u} \neq u$.
- Let $m = \min u = u(a)$ and let $M = \max u = u(b)$. From the previous observation, $M = 1 - m$. Then, (assume $b > a$, case $a \geqslant b$ is similar)
  $$\left\|u'\right\|_1 \geqslant \int_a^b \left|u'(t)\right| \mathrm{d}t \geqslant \int_a^b u'(t) = M - m = 1 - 2m.$$
  Also, since $m \leqslant 1 - m$, we must have $m \in [0, 1/2]$.

Motivating example

To summarize, we have shown that $u \in [m, 1-m]$ for some $m \in [0, 1/2]$, $u(1-t) = 1 - u(t)$, and

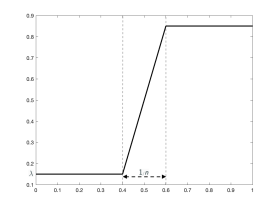
$$\mathcal{E}(u) \geqslant \lambda(M - m) + \int_0^{1/2} m^2 + \int_{1/2}^1 (1 - M)^2 = \lambda(1 - 2m) + m^2.$$

The RHS is minimal when $m = \lambda$ if $\lambda \leqslant 1/2$ and $m = 1/2$ if $\lambda \geqslant 1/2$. In the latter case, we see that $u \equiv 1/2$ achieves the minimum and is the unique minimizer.

Motivating example

Assume now that $\lambda < 1/2$. Then for any minimizer u, $\mathcal{E}(u) \geqslant \lambda(1-\lambda)$. Let us construct a minimizing sequence: For $n \geqslant 2$, define

$$u_n(t) = \begin{cases} \lambda & t \leqslant 1/2 - 1/n, \\ \frac{1}{2} + n(t-1/2)(1/2-\lambda) & |t-1/2| \leqslant 1/n, \\ 1-\lambda & t \geqslant 1/2 + 1/n. \end{cases}$$



- $\int_0^1 |u_n'| = \int_0^1 u_n' = 1 - 2\lambda$.
- $\mathcal{E}(u_n) \leqslant \lambda(1-2\lambda) + \left(1 - \frac{2}{n}\right)^2 \lambda^2 + \frac{2}{n} \to \lambda(1-\lambda)$ as $n \to \infty$. So $\inf_u \mathcal{E}(u) = \lambda(1-\lambda)$.

The L^1 limit of u_n is $u = \lambda \chi_{[0,1/2)} + (1-\lambda)\chi_{[1/2,1]}$, which is not in $W^{1,1}$. Note also that since $\int |u_n'| = 1 - 2\lambda$ for all n, it is natural to assume that $\int |u'|$ makes sense. A natural extension of the functional F is to define for $u \in L^1$:

$$F(u) = \inf \left\{ \lim_{n\to\infty} \int_0^1 \left|u_n'(t)\right| \mathrm{d}t \; ; \; u_n \to u \text{ in } L^1, \quad \lim_{n\to\infty} \int_0^1 |u_n'| < \infty \right\}.$$

This definition is consistent with the more standard definition of total variation.

Definition 10

Let $\Omega \subset \mathbb{R}^n$ be open and $u \in L^1(\Omega)$. Let $\mathcal{D}(\Omega; \mathbb{R}^n)$ be the following set of vector-valued test functions

$$\mathcal{D}(\Omega; \mathbb{R}^n) \stackrel{\text{def.}}{=} \left\{ \varphi \in C_c^\infty(\Omega; \mathbb{R}^n) \; ; \; \operatorname*{ess\,sup}_{x \in \Omega} \|\varphi(x)\|_2 \leqslant 1 \right\}.$$

The total variation of $u \in L^1(\Omega)$ is defined as

$$\mathrm{TV}(u) \stackrel{\text{def.}}{=} \sup_{\varphi \in \mathcal{D}(\Omega; \mathbb{R}^n)} \int_\Omega u(x) \mathrm{div} \varphi(x) \mathrm{d}x$$

Definition 11

The functions $u \in L^1(\Omega)$ with finite value of TV form a normed space called the space of functions of bounded variation defined as

$$\mathrm{BV}(\Omega) \overset{\text{def.}}{=} \left\{ u \in L^1(\Omega) \; ; \; \|u\|_{BV} \overset{\text{def.}}{=} \|u\|_{L^1} + \mathrm{TV}(u) < \infty \right\}.$$

- One can show that $\mathrm{BV}(\Omega)$ is a Banach space.
- For $u \in W^{1,1}(\Omega)$ with weak derivative ∇u, we have $TV(u) = \int_\Omega \|\nabla u\|_2 \, \mathrm{d}x$. So, $W^{1,1}(\Omega) \subset \mathrm{BV}(\Omega)$.
- However, $\mathrm{BV}(\Omega)$ is a larger space as it allows for discontinuous functions. For $C \subset \Omega$ with smooth boundary, we have $TV(1_C) = \mathrm{Per}(C)$.

Total variation regularisation

Lemma 12 (Lower semi-continuity properties of TV)

(i) *J is lower semicontinuous wrt weak convergence in L^p for $p \in [1, \infty)$.*

(ii) *J is convex.*

Proof.

Let

$$L_\varphi : u \mapsto -\int_\Omega u(x)\mathrm{div}\varphi(x)\mathrm{d}x.$$

If $u_n \rightharpoonup u$ in $L^p(\Omega)$, then $L_\varphi u_n \to L_\varphi u$. Note however that

$$L_\varphi u = \lim_{n\to\infty} L_\varphi u_n \leqslant \liminf_{n\to\infty} J(u_n).$$

Taking the supremum over all $\varphi \in C_c^\infty(\Omega, \mathbb{R}^N)$ with $\|\varphi\|_\infty \leqslant 1$ yields

$$J(u) \leqslant \liminf_{n\to\infty} J(u_n).$$

To see that J is convex, let $u_1, u_2 \in L^p(\Omega)$ and let $t \in [0, 1]$. Then,

$$L_\varphi(tu_1 + (1-t)u_2) = tL(u_1) + (1-t)L(u_2) \leqslant tJ(u_1) + (1-t)J(u_2).$$

Taking the supremum over all $\varphi \in C_c^\infty(\Omega, \mathbb{R}^N)$ with $\|\varphi\|_\infty \leqslant 1$ yields the

The space $BV(\Omega)$ can be compactly embedded into L^1. In contrast, note that no such compactness result exists for $W^{1,1}(\Omega)$ since one can in fact construct bounded sequences in $W^{1,1}(\Omega)$ which converge to elements of $BV(\Omega)$.

Theorem 13 (Rellich's compactness theorem)

Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary, and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of functions in $BV(\Omega)$ such that $\sup_n \|u_n\|_{BV} < \infty$. Then there exists $u \in BV(\Omega)$ and a subsequence $(u_{n_k})_{k\geqslant 1}$ such that $u_{n_k} \to u$ in $L^1(\Omega)$ as $k \to \infty$.

Theorem 3.1

Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary, let $u \in BV(\Omega)$. Then, there exists a sequence (u_n) of functions in $C^\infty(\Omega) \cap W^{1,1}(\Omega)$, such that

1. $u_n \to u$ in L^1.
2. $J(u_n) = \int_\Omega |\nabla u| \to J(u) = \int_\Omega |Du|$.

Total variation regularisation

Let $A : L^2(\Omega) \to L^2(\Omega)$ and consider

$$\min_{u \in L^2(\Omega)} \|Au - f_\delta\|_2^2 + \alpha \mathrm{TV}(u) \tag{3.1}$$

The case where $A = \mathrm{Id}$ is known as the Rudin-Osher-Fatemi model for denoising.

Let $A : L^2(\Omega) \to L^2(\Omega)$ and consider

$$\min_{u \in L^2(\Omega)} \|Au - f_\delta\|_2^2 + \alpha \mathrm{TV}(u) \tag{3.1}$$

The case where $A = \mathrm{Id}$ is known as the Rudin-Osher-Fatemi model for denoising.

Existence of solutions?

Let $A : L^2(\Omega) \to L^2(\Omega)$ and consider

$$\min_{u \in L^2(\Omega)} \|Au - f_\delta\|_2^2 + \alpha \mathrm{TV}(u) \tag{3.1}$$

The case where $A = \mathrm{Id}$ is known as the Rudin-Osher-Fatemi model for denoising.

Existence of solutions?

On L^2, lsc is ok, but the embedding of BV to L^2 is continuous but not compact.

Let $A : L^2(\Omega) \to L^2(\Omega)$ and consider

$$\min_{u \in L^2(\Omega)} \|Au - f_\delta\|_2^2 + \alpha \mathrm{TV}(u) \tag{3.1}$$

The case where $A = \mathrm{Id}$ is known as the Rudin-Osher-Fatemi model for denoising.

Existence of solutions?

On L^2, lsc is ok, but the embedding of BV to L^2 is continuous but not compact.

Theorem 3.2 (Poincaré inequality)

Let $\Omega \subset \mathbb{R}^N$. For $u \in BV(\Omega)$, let $m(u) = \frac{1}{|\Omega|} \int_\Omega u(x) \mathrm{d}x$. Then there exists $C > 0$ such that

$$\|u - m(u)\|_{L^p} \leqslant C \mathrm{TV}(u), \qquad \forall u \in BV(\Omega),$$

for all $p \in [1, N/(N-1)]$. In particular, this holds with $p = 2$ when $N = 2$.

Assume $A\chi_\Omega \neq 0$. Given u_n s.t. $\mathrm{TV}(u_n) + \|Au_n - f_\delta\|_2^2 \leqslant C$, $m(u_n)$ is also uniformly bounded:

- let $w_n = m(u_n)$ and $v_n = u_n - m(u_n)$. Then, $\int v_n = 0$ and $J(v_n) = J(u_n)$. So, by the Poincaré inequality, $\|v_n\|_2 \leqslant C'$.
- Observe now that $C \geqslant \|Au_n - f_\delta\|_2 \geqslant \|Aw_n\|_2 - \|Av_n - f_\delta\|_2$, and hence

$$C + \|A\| \, \|v_n\|_2 + \|f_\delta\|_2 \geqslant \|Aw_n\|_2 = \left| \int u_n \right| \frac{\|A\chi_\Omega\|_2}{|\Omega|}.$$

So, Poincaré inequality tells us that $\|u_n\|_2$ is uniformly bounded and we can extract a weakly convergent subsequence in L^2. Our convergent regularisation theorem guarantees convergence of u_δ weakly in L^2, but since BV is compactly embedded in L^1, we also know that u_δ converges strongly in L^1.

Subdifferential of the total variation functional

Recall that for each $u \in L^1(\Omega)$, $J(u) = \sup_{p \in \mathcal{K}} \int_\Omega u(x)p(x)\mathrm{d}x$ where

$$\mathcal{K} = \left\{ -\mathrm{div}\varphi \; ; \; \varphi \in C_c^\infty(\Omega; \mathbb{R}^N), \|\varphi\|_\infty \leqslant 1 \right\}.$$

However, if $u \in L^2(\Omega)$, then we in fact have:

$$J(u) = \sup_{p \in K} \int_\Omega u(x)p(x)\mathrm{d}x,$$

where

$$K = \left\{ -\mathrm{div}\varphi \; ; \; \varphi \in L^\infty(\Omega, \mathbb{R}^N), -\mathrm{div}\varphi \in L^2(\Omega), \varphi \cdot \eta_\Omega = 0 \right\}.$$

In the definition of K, $-\mathrm{div}\varphi \in L^2(\Omega)$ means that there exists $\gamma \in L^2(\Omega)$ such that

$$\int_\Omega \gamma u = \int z \cdot \nabla u, \qquad \forall u \in C_c^\infty(\Omega).$$

Since we are dealing with a one-homogeneous functional, we have $J(u) = J^{**}(u) = \sup_{p \in \partial J(0)} \langle p, \, u \rangle$, so

$$K = \left\{ p \in L^2(\Omega) \; ; \; \int_\Omega p(x)u(x)\mathrm{d}x \leqslant J(u), \quad \forall u \in L^2(\Omega) \right\}.$$

Moreover, $J(u) = \left\{ p \in L^2(\Omega) \; ; \; \langle p, \, u \rangle \leqslant J(u), \quad \forall u \in L^2 \right\}$.

Source condition example 1

Consider the case of TV denoising with $\mathcal{U} = \mathcal{V} = L^2(\Omega)$ and $C \subset \Omega$ has C^∞ boundary. Then

$$TV(1_C) = \mathrm{Per}(C) = \int_{\partial C} 1 = \int_{\partial C} \langle \eta_{\partial C}, \eta_{\partial C} \rangle.$$

Since $\eta_{\partial C} \in C^\infty(\partial C, \mathbb{R}^2)$ and $\|\eta_{\partial C}(x)\|_2 = 1$, we can extend to $\psi \in C_0^\infty(\Omega; \mathbb{R}^2)$ with $\sup_x \|\psi(x)\|_2 \leqslant 1$. Therefore, by the divergence theorem

$$TV(1_C) = \int_{\partial C} \langle \psi, \eta_{\partial C} \rangle = \int_C \mathrm{div}(\psi) = \langle \mathrm{div}(\psi), 1_C \rangle$$

and $\mathrm{div}(\psi) \in \partial TV(0)$. So, the source condition is satisfied.

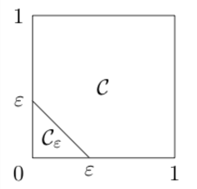
Source condition example 2

Suppose now that $C = [0,1]^2$ and suppose that $p_0 \in \partial TV(1_C) \subset L^2(\Omega)$. Then,

$$\langle p, 1_C \rangle = TV(1_C) = \mathrm{Per}(C) = 4.$$

Since $TV(u) \geqslant \langle p_0, u \rangle$ for all u,

$$TV(1_{C \setminus C_\varepsilon}) \geqslant \langle p_0, 1_{C \setminus C_\varepsilon} \rangle = \langle p_0, 1_C \rangle - \langle p_0, 1_{C_\varepsilon} \rangle$$



Source condition example 2

Suppose now that $C = [0,1]^2$ and suppose that $p_0 \in \partial TV(1_C) \subset L^2(\Omega)$. Then,

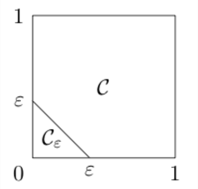
$$\langle p, 1_C \rangle = TV(1_C) = \mathrm{Per}(C) = 4.$$

Since $TV(u) \geqslant \langle p_0, u \rangle$ for all u,

$$4 - 2\varepsilon + \sqrt{2}\varepsilon = TV(1_{C \setminus C_\varepsilon}) \geqslant \langle p_0, 1_{C \setminus C_\varepsilon} \rangle = \langle p_0, 1_C \rangle - \langle p_0, 1_{C_\varepsilon} \rangle = 4 - \langle p_0, 1_{C_\varepsilon} \rangle$$

$$\frac{\varepsilon}{\sqrt{2}} \sqrt{\int_{C_\varepsilon} |p_0|^2} \geqslant |\langle p_0, 1_{C_\varepsilon} \rangle| \geqslant (2 - \sqrt{2})\varepsilon \implies \sqrt{\int_{C_\varepsilon} |p_0|^2} \geqslant \sqrt{2}(2 - \sqrt{2}).$$

Contradiction since $p_0 \in L^2$. Therefore $\partial J(1_C) = \emptyset$.



Theorem 3.3

Consider the setting of Theorem **??** *and* $v = -\mathrm{div} z$ *with* $\|z\|_\infty \leqslant 1$. *Let*

$$U_r \overset{\text{def.}}{=} \{x \in \Omega \; ; \; |z(x)| < r\} \, .$$

For each $r \in (0,1)$,

$$(1-r)\int_{U_r} |Du| \leqslant \frac{\delta^2}{2\lambda} + \frac{\lambda \|p\|_{L^2}^2}{2} + \delta \|p\|_{L^2} \, .$$

Proof.

$$\begin{aligned}
d &:= J(u) - J(u_0) - \langle v, \, u - u_0 \rangle \\
&= J(u) - J(u_0) + \langle \mathrm{div} z, \, u \rangle - \langle \mathrm{div} z, \, u_0 \rangle \\
&= J(u) + \langle \mathrm{div} z, \, u \rangle \qquad \text{since } J(u_0) = \langle -\mathrm{div} z, \, u_0 \rangle \\
&= J(u) - \int (z, Du) = J(u) - \int_{\Omega \setminus U_r} (z, Du) - \int_{U_r} (z, Du) \\
&\geqslant J(u) - \int_{\Omega \setminus U_r} |Du| - r \int_{U_r} |Du| \geqslant (1-r) \int_{U_r} |Du| \, .
\end{aligned}$$

Example

Let us consider the case of denoising. Then, the proof of Theorem **??** actually yields

$$\int_\Omega (u-u_0)^2 + |J(u)-J(u_0)| \leqslant C\delta(\|p\|_{L^2}+1) \tag{3.2}$$

provided $\lambda = \delta / \|p\|_{L^2}$. Let $B_R \subset \mathbb{R}^2$ be the ball of radius R with origin 0 and let $u_0 = \chi_{B_R}$. Then let $p = -\mathrm{div}(z)$ where z is defined by

$$z(x) = \frac{q\left(\Big||x|-R\Big|\right)}{|x|}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad q(s) = \max\{1-s/\varepsilon, 0\}.$$

(In polar coordinates (r,θ), we can write $z(r,\theta) = q(|r-R|)\binom{\cos(\theta)}{\sin(\theta)}$). One can show that $\|p\| = \mathcal{O}(\varepsilon^{-1/2})$. Then, by choosing $U = \{x \in \Omega\,;\ \mathrm{dist}(x, \partial B_R) \geqslant \varepsilon\}$, the minimizer u satisfies

$$\int_U |Du| \leqslant \mathcal{O}\left(\frac{\delta^2}{\lambda} + \frac{\lambda}{\varepsilon} + \frac{\delta}{\varepsilon}\right) = \mathcal{O}\left(\frac{\delta}{\sqrt{\varepsilon}}\right)$$

provided that $\lambda = \delta\sqrt{\varepsilon}$.

Combining with (3.2) yields

$$\int_{U^c} |Du| \geqslant J(u) - \int_U |Du| \geqslant 2\pi R - C\left(\delta + \frac{\delta}{\sqrt{\varepsilon}}\right).$$

Therefore, most of the total variation of u is concentrated around ∂B_R and the

Summary

To deal with the unboundedness of $A^\dagger$, regularisation aims to build a family of bounded operators R_α such that $R_\alpha f_\delta \to A^\dagger f$.

- The regularisation parameter α needs to be chosen appropriately to balance the data error $\|R_\alpha\|$ and the approximation error $\left\|R_\alpha f - A^\dagger f\right\|$.
- **Spectral regularisation** modifies the singular values of $A^\dagger$. E.g. by truncation.
- **Tikhonov** regularisation is a form of spectral regularisation, but does not require knowledge of the singular values. Corresponds to

$$\min_u \frac{1}{2} \|Au - f\|^2 + \mathcal{J}(u) \quad \text{where} \quad \mathcal{J}(u) = \|u\|^2 .$$

- Tikhonov regularisation is one of the early forms of variational regularisation, the choice of regulariser depends on our prior knowledge of the desired solutions. E.g. TV regularisation promotes sharp edges.
- For convex regularisers, convergence rates and solution properties can be analysed via the dual formulation and under the source condition.