

Inverse Problems

Variational regularisation

Clarice Poon
University of Bath

March 3, 2020

Variational regularisation

Background

Regularisation properties

Convergence rates

The dual perspective

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

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

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

$$\min_{u \in \mathcal{U}} \frac{1}{2} \|Au - f_\delta\|^2 + \frac{\alpha}{2} \|u\|^2. \quad (1.2)$$

Since this is a convex optimisation 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.

We will now study more general variational regularisers of the form

$$R_\alpha f_\delta \in \operatorname{argmin}_{u \in \mathcal{U}} \frac{1}{2} \|Au - f_\delta\|_{\mathcal{V}}^2 + \alpha \mathcal{J}(u). \quad (1.3)$$

where

- $A : \mathcal{U} \rightarrow \mathcal{V}$ is a bounded linear operator between a Banach spaces \mathcal{U} and a Hilbert space \mathcal{V} .
- $\mathcal{J} : \mathcal{U} \rightarrow [0, \infty]$.
- $f_\delta \in \mathcal{V}$ satisfies $\|Au^\dagger - f_\delta\|_{\mathcal{V}} \leq \delta$.

Example: smoothing regularisers

Let $\mathcal{J}(u) = \|Lu\|_{\mathcal{Z}}$ where $L : \mathcal{U} \rightarrow \mathcal{Z}$ is a linear (possibly unbounded) operator . Popular choices include differential operators, e.g. $L = \nabla$, $\mathcal{U} = W^{1,2}(\Omega)$, $\mathcal{Z} = L^2(\Omega)$.

Example: smoothing regularisers

Let $\mathcal{J}(u) = \|Lu\|_{\mathcal{Z}}$ where $L : \mathcal{U} \rightarrow \mathcal{Z}$ is a linear (possibly unbounded) operator . Popular choices include differential operators, e.g. $L = \nabla$, $\mathcal{U} = W^{1,2}(\Omega)$, $\mathcal{Z} = L^2(\Omega)$.

For $\mathcal{J}(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2$, u is a minimizer if and only if

$$A^*Au - A^*f - \alpha\Delta u = 0,$$

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

Example: smoothing regularisers

Let $\mathcal{J}(u) = \|Lu\|_{\mathcal{Z}}$ where $L : \mathcal{U} \rightarrow \mathcal{Z}$ is a linear (possibly unbounded) operator . Popular choices include differential operators, e.g. $L = \nabla$, $\mathcal{U} = W^{1,2}(\Omega)$, $\mathcal{Z} = L^2(\Omega)$.

For $\mathcal{J}(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2$, u is a minimizer if and only if

$$A^*Au - A^*f - \alpha\Delta u = 0,$$

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 f , so noise is removed.

Example: smoothing regularisers

Let $\mathcal{J}(u) = \|Lu\|_{\mathcal{Z}}$ where $L : \mathcal{U} \rightarrow \mathcal{Z}$ is a linear (possibly unbounded) operator . Popular choices include differential operators, e.g. $L = \nabla$, $\mathcal{U} = W^{1,2}(\Omega)$, $\mathcal{Z} = L^2(\Omega)$.

For $\mathcal{J}(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2$, u is a minimizer if and only if

$$A^*Au - A^*f - \alpha\Delta u = 0,$$

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 f , so noise is removed.
- For imaging applications, leads to oversmooth reconstructions as Δ has very strong isotropic smoothing properties.

Example: Lasso

Consider $\mathcal{U} = \mathcal{V} = \ell_2(\mathbb{N})$ and $\mathcal{J}(u) = \begin{cases} \|u\|_1 & u \in \ell_1(\mathbb{N}) \\ +\infty & u \in \ell_2(\mathbb{N}) \setminus \ell_1(\mathbb{N}) \end{cases}$.

The problem

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

is called the lasso in statistics and can be shown to promote sparse solutions.

One can also consider $\mathcal{J}(u) = \|Wu\|_1$ where $W : \ell_2(\mathbb{N}) \rightarrow \ell_2(\mathbb{N})$. For example, W is some wavelet transform.

Example: Lasso

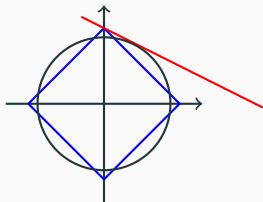
Consider $\mathcal{U} = \mathcal{V} = \ell_2(\mathbb{N})$ and $\mathcal{J}(u) = \begin{cases} \|u\|_1 & u \in \ell_1(\mathbb{N}) \\ +\infty & u \in \ell_2(\mathbb{N}) \setminus \ell_1(\mathbb{N}) \end{cases}$.

The problem

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

is called the lasso in statistics and can be shown to promote sparse solutions.

One can also consider $\mathcal{J}(u) = \|Wu\|_1$ where $W : \ell_2(\mathbb{N}) \rightarrow \ell_2(\mathbb{N})$. For example, W is some wavelet transform.



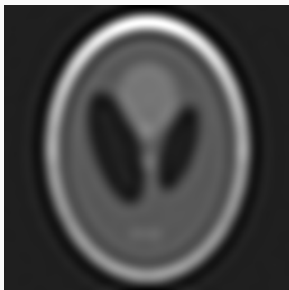
Consider $\langle u, a \rangle = f$ where $u \in \mathbb{R}^2$ is unknown, $a \in \mathbb{R}^2$ and $f \in \mathbb{R}$. Solutions are along the red line. The solution of smallest ℓ_1 norm will be 1-sparse, whereas the solution of smallest ℓ_2 norm is 2-sparse.

Example: Total variation

Instead of $\mathcal{J}(u) = \int_{\Omega} |\nabla u|^2$, one could consider $\mathcal{J}(u) = \int_{\Omega} |\nabla u|$.

Deblurring example:

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



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



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

Example: Total variation

The use of $\int_{\Omega} |\nabla u|^2$ leads to smooth solutions, the point of $\int_{\Omega} |\nabla u|$ is that this makes sense not only for $u \in W^{1,1}(\Omega)$ but also for functions of bounded variation.

Given $u \in L^1(\Omega)$ for $\Omega \subset \mathbb{R}^d$, define

$$\mathrm{TV}(u) \stackrel{\text{def.}}{=} \sup \left\{ \langle u, \operatorname{div} \varphi \rangle ; \varphi \in C_c^\infty(\Omega; \mathbb{R}^d), \sup_{\omega \in \Omega} \|\varphi(\omega)\|_2 \leq 1 \right\}.$$

Let $\|u\|_{BV} \stackrel{\text{def.}}{=} \|u\|_{L^1} + \mathrm{TV}(u)$, and the space of bounded variations $\{u \in L^1 ; \mathrm{TV}(u) < \infty\}$ is a Banach space with norm $\|\cdot\|_{BV}$.

Contains $W^{1,1}(\Omega)$ and also discontinuous functions such as χ_C where $C \subset \Omega$ has Lipschitz boundary, in which case, $\mathrm{TV}(\chi_C) = \operatorname{Per}(C)$.

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

1. A **data model**: $f = Au_0 + n$ where $u_0 \in \mathbb{R}^N$ is the underlying object to be recovered, A 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.

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

$$P(u|f)P(f) = P(f|u)P(u).$$

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

$$P(u|f)P(f) = P(f|u)P(u).$$

Choosing $P(f|u) = \exp\left(-\frac{1}{\sigma^2} \|f - Au\|_2^2\right)$:

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

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

$$P(u|f)P(f) = P(f|u)P(u).$$

Choosing $P(f|u) = \exp\left(-\frac{1}{\sigma^2} \|f - Au\|_2^2\right)$:

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

The maximum a posteriori (MAP) reconstruction is:

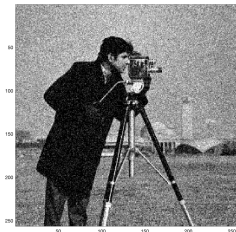
$$u^* \in \operatorname{argmax}_u P(u|f).$$

Equivalently, $u^* \in \operatorname{argmin}_u p(u) + \frac{1}{\sigma^2} \|f - Au\|_2^2.$

Bayesian viewpoint of variational methods

Other choices of noise distributions:

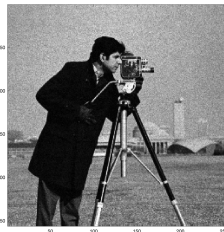
- Additive Laplace noise $e^{-\frac{1}{\sigma^2} \|f - Au\|_1}$ with corresponding data fidelity term $\|Au - f\|_1$
- Poisson noise $\prod_{i,j} \frac{f_{i,j}^{u_{i,j}}}{f_{i,j}!} e^{-u_{i,j}}$ with data fidelity term $\int u - f \log(u)$.



Gaussian



Impulse



Poisson

Figure 1: Adding different noise using Matlab's `imnoise` function

We now study regularisers of the form

$$R_\alpha(f) \in \operatorname{argmin}_u \alpha \mathcal{J}(u) + \frac{1}{2} \|f - Au\|_2^2.$$

Usual questions:

- Given $f = Au^\dagger$, do we have convergence $R_\alpha(f) \rightarrow u^\dagger$?
- Do we have convergent regularisers?
- Convergence rates?

Variational regularisation

Background

Regularisation properties

Convergence rates

The dual perspective

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

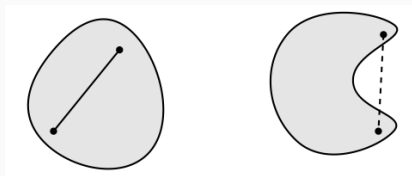
- Useful to model constraints. E.g. if $E : [-1, \infty) \rightarrow \mathbb{R}^2$ maps $x \mapsto x^2$, consider instead $\bar{E} : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ defined by $\bar{E}(x) = E(x)$ for $x \in [-1, \infty)$ and $\bar{E}(x) = +\infty$ otherwise. No need to worry if $E(x + y)$ is 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 $\iota_C \stackrel{\text{def.}}{=} \begin{cases} 1 & x \in C \\ +\infty & x \notin C \end{cases}$

So, we can write $\min_{u \in C} E(u) = \min_{u \in \mathcal{U}} E(u) + \iota_C(u)$.

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

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



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

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

It is called strictly convex if the inequality is strict.

Banach spaces are complete, normed 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

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

Banach spaces are complete, normed vector spaces.

Bi-dual

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

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

$E : \mathcal{U} \rightarrow \mathcal{U}^{**}$ is well defined and is a continuous linear isometry. If E is surjective, 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} .

Banach spaces are complete, normed vector spaces.

Adjoint

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

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

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 exist $u \in \ell^2$ such that $\|e_j - u\| \rightarrow 0$.

Weak and weak-* convergence

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 exist $u \in \ell^2$ such that $\|e_j - u\| \rightarrow 0$.

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 $\langle p, u_k \rangle \rightarrow \langle p, u \rangle$.

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 $\langle p_k, u \rangle \rightarrow \langle p, u \rangle$ for all $u \in \mathcal{U}$.

Weak and weak-* convergence

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 exist $u \in \ell^2$ such that $\|e_j - u\| \rightarrow 0$.

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 $\langle p, u_k \rangle \rightarrow \langle p, u \rangle$.

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 $\langle p_k, u \rangle \rightarrow \langle p, u \rangle$ for all $u \in \mathcal{U}$.

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

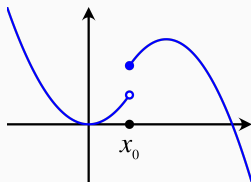
Lower semi-continuity

One useful property is the notion of sequential lower semicontinuity:

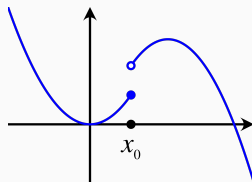
Let \mathcal{X} be a Banach space with topology $\tau_{\mathcal{X}}$. The functional $E : \mathcal{X} \rightarrow [-\infty, \infty]$ is said to be sequentially lower semi-continuous with respect to $\tau_{\mathcal{X}}$ at $u \in \mathcal{X}$ if

$$E(x) \leq \liminf_{j \rightarrow \infty} E(x_j)$$

for all sequences $\{x_j\}_j \subset \mathcal{X}$ with $x_j \rightarrow x$ in the topology $\tau_{\mathcal{X}}$ of \mathcal{X} .



Not lsc



lsc

" $E(x_0)$ is a good lower bound for function values near x_0 "

Example: Lower semi-continuity

Let \mathcal{U} be any normed space with norm $\|\cdot\|_{\mathcal{U}}$, then $E(u) = \|u\|_{\mathcal{U}}$ is lower semicontinuous with respect to the weak topology:

Example: Lower semi-continuity

Let \mathcal{U} be any normed space with norm $\|\cdot\|_{\mathcal{U}}$, then $E(u) = \|u\|_{\mathcal{U}}$ is lower semicontinuous with respect to the weak topology:

Idea: For fixed $u \in \mathcal{U}$, Hahn Banach Theorem says we can construct an element of $f \in \mathcal{U}^*$ such that $f(u) = \|u\|$ and $\|f\| = 1$.

Example: Lower semi-continuity

Let \mathcal{U} be any normed space with norm $\|\cdot\|_{\mathcal{U}}$, then $E(u) = \|u\|_{\mathcal{U}}$ is lower semicontinuous with respect to the weak topology:

Idea: For fixed $u \in \mathcal{U}$, Hahn Banach Theorem says we can construct an element of $f \in \mathcal{U}^*$ such that $f(u) = \|u\|$ and $\|f\| = 1$.

Proof: Let $u^j \rightarrow u$ weakly, by the Hahn-Banach theorem, there exists an element $f \in \mathcal{U}^*$ such that $f(u) = \|u\|_{\mathcal{U}}$ and $\|f\| = 1$. Therefore,

$$\|u\|_{\mathcal{U}} = f(u) = \lim_j f(u^j) \leq \liminf_j \|u^j\|_{\mathcal{U}}.$$

Example: Lower semi-continuity

The functional $\|\cdot\|_1 : \ell^2 \rightarrow [0, \infty]$ is lower semi-continuous with respect to ℓ_2 convergence.

Example: Lower semi-continuity

The functional $\|\cdot\|_1 : \ell^2 \rightarrow [0, \infty]$ is lower semi-continuous with respect to ℓ_2 convergence.

Idea: Convergence in ℓ_2 implies that each entry of the sequence converges. So, we can just apply Fatou's lemma.

Example: Lower semi-continuity

The functional $\|\cdot\|_1 : \ell^2 \rightarrow [0, \infty]$ is lower semi-continuous with respect to ℓ_2 convergence.

Idea: Convergence in ℓ_2 implies that each entry of the sequence converges. So, we can just apply Fatou's lemma.

Proof: Given any $\{u^j\} \subset \ell_2$ with $u^j \rightarrow u$ in ℓ_2 , we have

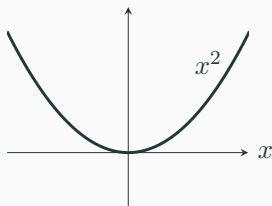
$$u_k^j = \langle e_k, u^j \rangle \rightarrow \langle e_k, u \rangle = u_k.$$

So, by Fatou's lemma

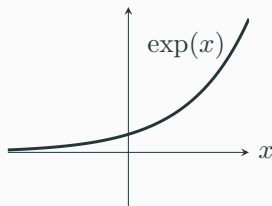
$$\|u\|_1 = \sum_k \lim_{j \rightarrow \infty} |u_k^j| \leq \liminf_{j \rightarrow \infty} \sum_k |u_k^j| = \liminf_{j \rightarrow \infty} \|u^j\|_1.$$

Minimising functionals

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



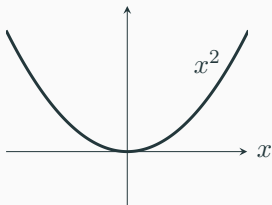
Coercive



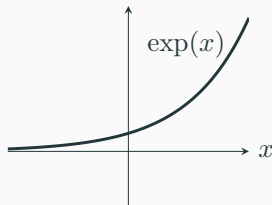
Not coercive.

Minimising functionals

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



Coercive



Not coercive.

Coercivity is **sufficient** to ensure boundedness of minimising sequences:

Lemma 2.1

Let $E : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ 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 $E(u_j) \rightarrow \inf_u E(u)$ and all minimising sequences are bounded.

Theorem 2.2 (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 $E : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ be proper coercive and $\tau_{\mathcal{U}}$ -l.s.c, and bounded from below. Then E has a minimiser.

Theorem 2.2 (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 $E : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ be proper coercive and $\tau_{\mathcal{U}}$ -l.s.c, and bounded from below. Then E has a minimiser.

Idea: Bounded sequences have convergent subsequences. So we take a minimising sequences, and show its limit is a minimiser.

Theorem 2.2 (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 $E : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ be proper coercive and $\tau_{\mathcal{U}}$ -l.s.c, and bounded from below. Then E has a minimiser.

Idea: Bounded sequences have convergent subsequences. So we take a minimising sequences, and show its limit is 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_{j_k} and $u_* \in \mathcal{U}$ which converges $\tau_{\mathcal{U}}$ to u_* .
- Due to $\tau_{\mathcal{U}}$ -lsc, we have $E(u^*) \leq \liminf_{k \rightarrow \infty} E(u_{j_k}) = \inf_u E(u) > \infty$. 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

$$E(u) \leq E\left(\frac{1}{2}u + \frac{1}{2}v\right) < \frac{1}{2}E(u) + \frac{1}{2}E(v) \leq E(u)$$

which is a contradiction. Not however that strict convexity is not necessary for uniqueness of minimisers (e.g. think for $f(x) = |x|$).

Variational regularisation

Background

Regularisation properties

Convergence rates

The dual perspective

We now study the properties of

$$R_\alpha f \in \operatorname{argmin}_{u \in \mathcal{U}} \Phi_{\alpha, f}(u) \stackrel{\text{def.}}{=} \frac{1}{2} \|Au - f\|_{\mathcal{V}}^2 + \alpha \mathcal{J}(u)$$

as a convergent regularisation for

$$Au = f \tag{3.1}$$

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

1. When do minimisers exist? (i.e. well-posedness of the regularised problem)
2. Is $R_\alpha : \mathcal{V} \rightarrow \mathcal{U}$ continuous?
3. What is the equivalent notion of a minimal norm solution here?
4. How to choose $\alpha(\delta)$ to guarantee the convergence of the minimisers to an appropriated generalised solution?

1. Existence of minimisers

Theorem 1

Let \mathcal{U} be a Banach space and let \mathcal{V} be a Hilbert space with topologies $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ respectively. Let $\|\cdot\|_{\mathcal{V}}$ be $\tau_{\mathcal{V}}$ -lsc. Assume that

- (i) $A : \mathcal{U} \rightarrow \mathcal{V}$ is $\tau_{\mathcal{U}} \rightarrow \tau_{\mathcal{V}}$ continuous.
- (ii) $\mathcal{J} : \mathcal{U} \rightarrow (0, +\infty]$ is proper, $\tau_{\mathcal{U}}$ -lsc and its non-empty sublevel-sets $\{u \in \mathcal{U} ; \mathcal{J}(u) \leq C\}$ are $\tau_{\mathcal{U}}$ -sequentially compact

Then,

- (i') for any fixed $\alpha > 0$ and $f \in \mathcal{V}$, there exists a minimiser of $u^{\alpha} \in \operatorname{argmin}_u \frac{1}{2} \|Au - f\|_{\mathcal{V}}^2 + \alpha \mathcal{J}(u)$.
- (ii') If A is injective or \mathcal{J} is strictly convex, then u^{α} is unique.

1. Existence of minimisers

Theorem 1

Let \mathcal{U} be a Banach space and let \mathcal{V} be a Hilbert space with topologies $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ respectively. Let $\|\cdot\|_{\mathcal{V}}$ be $\tau_{\mathcal{V}}$ -lsc. Assume that

- (i) $A : \mathcal{U} \rightarrow \mathcal{V}$ is $\tau_{\mathcal{U}} \rightarrow \tau_{\mathcal{V}}$ continuous.
- (ii) $\mathcal{J} : \mathcal{U} \rightarrow (0, +\infty]$ is proper, $\tau_{\mathcal{U}}$ -lsc and its non-empty sublevel-sets $\{u \in \mathcal{U} ; \mathcal{J}(u) \leq C\}$ are $\tau_{\mathcal{U}}$ -sequentially compact

Then,

- (i') for any fixed $\alpha > 0$ and $f \in \mathcal{V}$, there exists a minimiser of $u^{\alpha} \in \operatorname{argmin}_u \frac{1}{2} \|Au - f\|_{\mathcal{V}}^2 + \alpha \mathcal{J}(u)$.
- (ii') If A is injective or \mathcal{J} is strictly convex, then u^{α} is unique.

Idea: the direct method of Calculus. Take a minimising sequence, show it has a limit (from a subsequence), then use l.s.c. properties to conclude it is a minimiser.

1. Existence of minimisers

Since $\Phi_{\alpha,f}(u) \geq 0$, there exists a minimising sequence u_j so that

$$\lim_{j \rightarrow \infty} \Phi_{\alpha,f}(u_j) = \inf_{u \in \mathcal{U}} \Phi_{\alpha,f}(u) \stackrel{\text{def.}}{=} L.$$

In particular, $J(u_j)$ is uniformly bounded. Since the level sets of \mathcal{J} are $\tau_{\mathcal{U}}$ sequentially compact, there exists a subsequence u_{j_k} which converges $\tau_{\mathcal{U}}$ to some $u \in \mathcal{U}$.

By continuity of A , Au_{j_k} converges to Au in $\tau_{\mathcal{V}}$. By lsc properties of \mathcal{J} and $\|\cdot\|_{\mathcal{V}}$, we have

$$\Phi_{\alpha,f}(u) \leq \liminf_{k \rightarrow \infty} \Phi_{\alpha,f}(u_{j_k}) \leq L.$$

Therefore, u is a minimiser.

Finally, we saw that the minimum is unique if $\Phi_{\alpha,f}$ is strictly convex. Note that $u \mapsto \|Au - f\|_{\mathcal{V}}$ is strictly convex if and only if A is injective (exercise!).

2. Variational regularisers are continuous

Theorem 2

Fix $\alpha > 0$. Under the assumptions of Theorem 4, assume also

- either A is injective or \mathcal{J} is strictly convex.
- norm convergence in \mathcal{V} implies convergence in $\tau_{\mathcal{V}}$.

Then, given $f_j \rightarrow f$ in \mathcal{V} , $u_j \stackrel{\text{def.}}{=} R_{\alpha} f_j$ exists and is unique, and u_j converges to $u \stackrel{\text{def.}}{=} R_{\alpha} f$ in $\tau_{\mathcal{U}}$. Moreover, $\mathcal{J}(u_j) \rightarrow \mathcal{J}(u)$.

2. Variational regularisers are continuous

Theorem 2

Fix $\alpha > 0$. Under the assumptions of Theorem 4, assume also

- either A is injective or \mathcal{J} is strictly convex.
- norm convergence in \mathcal{V} implies convergence in $\tau_{\mathcal{V}}$.

Then, given $f_j \rightarrow f$ in \mathcal{V} , $u_j \stackrel{\text{def.}}{=} R_{\alpha} f_j$ exists and is unique, and u_j converges to $u \stackrel{\text{def.}}{=} R_{\alpha} f$ in $\tau_{\mathcal{U}}$. Moreover, $\mathcal{J}(u_j) \rightarrow \mathcal{J}(u)$.

Idea: As before,

1. we first show that $\{\Phi_{\alpha,f}(u_j)\}_j$ is bounded
2. This lets us extract a convergent subsequence with limit \hat{u} .
3. Finally, show that \hat{u} minimises $\Phi_{\alpha,f}$.

2. Variational regularisers are continuous (proof)

Step 1: show that $\Phi_{\alpha,f}(u_j)$ is bounded: First observe that

(a) $\|f + g\|_{\mathcal{V}}^2 \leq 2\|f\|_{\mathcal{V}}^2 + 2\|g\|_{\mathcal{V}}^2$ for all $f, g \in \mathcal{V}$.

(b) From (a), we have

$$\Phi_{\alpha,f}(u) \leq \|Au - g\|^2 + \|g - f\|^2 + 2\alpha\mathcal{J}(u) \leq 2\Phi_{\alpha,g}(u) + \|f - g\|^2.$$

Now, since \mathcal{J} is proper, there exists \tilde{u} such that $\Phi_{\alpha,f}(\tilde{u}) < \infty$

$$\Phi_{\alpha,f}(u_j) \leq 2\Phi_{\alpha,f_j}(u_j) + \|f - f_j\|_{\mathcal{V}}^2 \leq 2\Phi_{\alpha,f_j}(\tilde{u}) + \|f - f_j\|_{\mathcal{V}}^2$$

Step 2, Extract a subsequence which converges to \hat{u} : By compactness of the sublevel sets of \mathcal{J} , there exists a subsequence u_{j_k} which converges $\tau_{\mathcal{U}}$ to some $\hat{u} \in \mathcal{U}$.

2. Variational regularisers are continuous (proof)

Step 3, $\hat{u} = u$: By continuity of A , lsc of $\|\cdot\|_{\mathcal{V}}$ and lsc of \mathcal{J} , we have

$$\Phi_{\alpha,f}(\hat{u}) \leq \liminf_k \Phi_{\alpha,f_{j_k}}(u_{j_k}) \leq \liminf \Phi_{\alpha,f_{j_k}}(u) = \Phi_{\alpha,f}(u).$$

By uniqueness of minimisers, $\hat{u} = u$

Step 4, the entire sequence converges: Repeat this for any subsequence of $\{u_j\}$ to see that all subsequences have a subsequence which converge to u . Therefore, the entire sequence u_j converges to u in $\tau_{\mathcal{U}}$.

For the last statement, We see from Step 3 that $\Phi_{\alpha,f_j}(u_j) \rightarrow \Phi_{\alpha,f}(u)$. So,

$$\begin{aligned} \limsup_{j \rightarrow \infty} \alpha \mathcal{J}(u_j) &= \limsup_{j \rightarrow \infty} \Phi_{\alpha,f_j}(u_j) - \frac{1}{2} \|Au_j - f_j\|^2 \\ &= \Phi_{\alpha,f}(u) - \liminf_{j \rightarrow \infty} \|Au_j - f_j\|^2 \leq \Phi_{\alpha,f}(u) - \|Au - f\|^2 \\ &= \alpha \mathcal{J}(u) \leq \liminf_{j \rightarrow \infty} \alpha \mathcal{J}(u_j). \end{aligned}$$

3. \mathcal{J} -minimising solutions

Definition 3 (\mathcal{J} -minimising solutions)

Let

- $u_{\mathcal{J}}^{\dagger} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}$ and
- $\mathcal{J}(u_{\mathcal{J}}^{\dagger}) \leq \mathcal{J}(\tilde{u})$ for all $\tilde{u} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|$.

Then, $u_{\mathcal{J}}^{\dagger}$ is called a \mathcal{J} -minimising solution of the problem $Au = f$.

3. \mathcal{J} -minimising solutions

Definition 3 (\mathcal{J} -minimising solutions)

Let

- $u_{\mathcal{J}}^{\dagger} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}$ and
- $\mathcal{J}(u_{\mathcal{J}}^{\dagger}) \leq \mathcal{J}(\tilde{u})$ for all $\tilde{u} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|$.

Then, $u_{\mathcal{J}}^{\dagger}$ is called a \mathcal{J} -minimising solution of the problem $Au = f$.

3. \mathcal{J} -minimising solutions

Definition 3 (\mathcal{J} -minimising solutions)

Let

- $u_{\mathcal{J}}^{\dagger} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}$ and
- $\mathcal{J}(u_{\mathcal{J}}^{\dagger}) \leq \mathcal{J}(\tilde{u})$ for all $\tilde{u} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|$.

Then, $u_{\mathcal{J}}^{\dagger}$ is called a \mathcal{J} -minimising solution of the problem $Au = f$.

- As \mathcal{V} is a Hilbert space, $\mathbb{L}_f \stackrel{\text{def.}}{=} \{v ; v \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}\}$ is non-empty if and only if $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$.

3. \mathcal{J} -minimising solutions

Definition 3 (\mathcal{J} -minimising solutions)

Let

- $u_{\mathcal{J}}^{\dagger} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}$ and
- $\mathcal{J}(u_{\mathcal{J}}^{\dagger}) \leq \mathcal{J}(\tilde{u})$ for all $\tilde{u} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|$.

Then, $u_{\mathcal{J}}^{\dagger}$ is called a \mathcal{J} -minimising solution of the problem $Au = f$.

- As \mathcal{V} is a Hilbert space, $\mathbb{L}_f \stackrel{\text{def.}}{=} \{v ; v \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}\}$ is non-empty if and only if $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$. When it is no ambiguity, we write $\mathbb{L} = \mathbb{L}_f$.

3. \mathcal{J} -minimising solutions

Definition 3 (\mathcal{J} -minimising solutions)

Let

- $u_{\mathcal{J}}^{\dagger} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}$ and
- $\mathcal{J}(u_{\mathcal{J}}^{\dagger}) \leq \mathcal{J}(\tilde{u})$ for all $\tilde{u} \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|$.

Then, $u_{\mathcal{J}}^{\dagger}$ is called a \mathcal{J} -minimising solution of the problem $Au = f$.

- As \mathcal{V} is a Hilbert space, $\mathbb{L}_f \stackrel{\text{def.}}{=} \{v ; v \in \operatorname{argmin}_{u \in \mathcal{U}} \|Au - f\|_{\mathcal{V}}\}$ is non-empty if and only if $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$. When it is no ambiguity, we write $\mathbb{L} = \mathbb{L}_f$.
- We next establish existence under appropriate compactness and continuity assumptions. Note however: even when there is existence, in general, there is no uniqueness.

3. Existence of a \mathcal{J} -minimising solution

Theorem 4

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 $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$ and \mathbb{L} has an element with finite \mathcal{J} -value. Assume also that

- (i) $A : \mathcal{U} \rightarrow \mathcal{V}$ is $\tau_{\mathcal{U}} \rightarrow \tau_{\mathcal{V}}$ continuous.*
- (ii) $\mathcal{J} : \mathcal{U} \rightarrow (0, +\infty]$ is proper, $\tau_{\mathcal{U}}$ -lsc and its non-empty sublevel-sets $\{u \in \mathcal{U} ; \mathcal{J}(u) \leq C\}$ are $\tau_{\mathcal{U}}$ -sequentially compact*

Then, there exists a \mathcal{J} -minimising solution $u_{\mathcal{J}}^{\dagger}$.

3. Existence of a \mathcal{J} -minimising solution

Theorem 4

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 $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp}$ and \mathbb{L} has an element with finite \mathcal{J} -value. Assume also that

- (i) $A : \mathcal{U} \rightarrow \mathcal{V}$ is $\tau_{\mathcal{U}} \rightarrow \tau_{\mathcal{V}}$ continuous.
- (ii) $\mathcal{J} : \mathcal{U} \rightarrow (0, +\infty]$ is proper, $\tau_{\mathcal{U}}$ -lsc and its non-empty sublevel-sets $\{u \in \mathcal{U} ; \mathcal{J}(u) \leq C\}$ are $\tau_{\mathcal{U}}$ -sequentially compact

Then, there exists a \mathcal{J} -minimising solution $u_{\mathcal{J}}^{\dagger}$.

Proof: Consider $\inf_{u \in \mathbb{L}} \mathcal{J}(u)$. Note that \mathbb{L} is nonempty by assumption.

- Since $\mathcal{J} \geq 0$, 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_* . Moreover, continuity of A means Au_{n_k} converges to Au_* in $\tau_{\mathcal{V}}$.
- $u_* \in \mathbb{L}$ since $\|Au_* - f\| \leq \liminf_{k \rightarrow \infty} \|Au_{n_k} - f\| \leq \inf_u \|Au - f\|$.
- u_* is a minimiser as \mathcal{J} is $\tau_{\mathcal{U}}$ -lsc: $\inf_{u \in \mathbb{L}} \mathcal{J}(u) = \liminf_k \mathcal{J}(u_{n_k}) \geq \mathcal{J}(u_*)$.

4. Convergent regularisation

Theorem 5

Under the assumptions of Theorem 4, if $\alpha = \alpha(\delta)$ is such that $\alpha(\delta) \rightarrow 0$ and $\delta^2/\alpha(\delta) \rightarrow 0$ as $\delta \rightarrow 0$, then $u_\delta \stackrel{\text{def.}}{=} u_\delta^{\alpha(\delta)}$ converges (up to a subsequence) $\tau_{\mathcal{U}}$ to a \mathcal{J} minimising solution $u_{\mathcal{J}}^\dagger$ and $\mathcal{J}(u_\delta) \rightarrow \mathcal{J}(u_{\mathcal{J}}^\dagger)$.

4. Convergent regularisation

Theorem 5

Under the assumptions of Theorem 4, if $\alpha = \alpha(\delta)$ is such that $\alpha(\delta) \rightarrow 0$ and $\delta^2/\alpha(\delta) \rightarrow 0$ as $\delta \rightarrow 0$, then $u_\delta \stackrel{\text{def.}}{=} u_\delta^{\alpha(\delta)}$ converges (up to a subsequence) $\tau_{\mathcal{U}}$ to a \mathcal{J} minimising solution $u_{\mathcal{J}}^\dagger$ and $\mathcal{J}(u_\delta) \rightarrow \mathcal{J}(u_{\mathcal{J}}^\dagger)$.

Idea: Show that $\{\mathcal{J}(u_\delta)\}_\delta$ is bounded. Then, use compactness and lsc properties to deduce that it has a limit (up to subsequence) which is a \mathcal{J} -minimising solution.

4. Convergent regularisation

Theorem 5

Under the assumptions of Theorem 4, if $\alpha = \alpha(\delta)$ is such that $\alpha(\delta) \rightarrow 0$ and $\delta^2/\alpha(\delta) \rightarrow 0$ as $\delta \rightarrow 0$, then $u_\delta \stackrel{\text{def.}}{=} u_\delta^{\alpha(\delta)}$ converges (up to a subsequence) $\tau_{\mathcal{U}}$ to a \mathcal{J} minimising solution $u_{\mathcal{J}}^\dagger$ and $\mathcal{J}(u_\delta) \rightarrow \mathcal{J}(u_{\mathcal{J}}^\dagger)$.

- Since u_δ is a minimiser:

$\|Au_\delta - f_\delta\|^2 + \alpha(\delta)\mathcal{J}(u_\delta) \leq \frac{1}{2} \|Au_{\mathcal{J}}^\dagger - f_\delta\|^2 + \alpha(\delta)\mathcal{J}(u_{\mathcal{J}}^\dagger)$. This implies that $\mathcal{J}(u_\delta) \leq \mathcal{J}(u_{\mathcal{J}}^\dagger) + \frac{\delta^2}{2\alpha(\delta)}$.

- by compactness of the sublevel sets of \mathcal{J} , up to a subsequence u_{δ_n} converges to u_* as $\delta_n \rightarrow 0$. By continuity of A , $Au_{\delta_n} \xrightarrow{\tau_{\mathcal{V}}} Au_*$.
- $u_* \in \mathbb{L}_f$ follows by lsc of $\|\cdot\|_{\mathcal{V}}$ wrt $\tau_{\mathcal{V}}$ and by minimality of u_{δ_n} :

$$\begin{aligned} \frac{1}{2} \|Au_* - f\|^2 &\leq \liminf_{n \rightarrow \infty} \frac{1}{2} \|Au_{\delta_n} - f_{\delta_n}\|^2 \leq \liminf_{n \rightarrow \infty} \frac{1}{2} \|Au_{\delta_n} - f_{\delta_n}\|^2 + \alpha(\delta_n)\mathcal{J}(u_{\delta_n}) \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{2} \|Au_{\mathcal{J}}^\dagger - f_{\delta_n}\|^2 + \alpha(\delta_n)\mathcal{J}(u_{\mathcal{J}}^\dagger) = \inf_u \|Au - f\|. \end{aligned}$$

- Finally

$$\mathcal{J}(u_*) \leq \liminf_{n \rightarrow \infty} \mathcal{J}(u_{\delta_n}) \leq \liminf_{n \rightarrow \infty} \mathcal{J}(u_{\mathcal{J}}^\dagger) + \frac{\delta_n^2}{2\alpha(\delta_n)} = \mathcal{J}(u_{\mathcal{J}}^\dagger).$$

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

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

- Lower semicontinuity: \mathcal{J} is weakly-lsc

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

- Lower semicontinuity: \mathcal{J} is weakly-lsc
- Compactness of sublevel sets: bounded sequences have weakly convergent subsequences.

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

- Lower semicontinuity: \mathcal{J} is weakly-lsc
- Compactness of sublevel sets: bounded sequences have weakly convergent subsequences.

So, Theorem 5 holds with weak convergence.

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

- Lower semicontinuity: \mathcal{J} is weakly-lsc
- Compactness of sublevel sets: bounded sequences have weakly convergent subsequences.

So, Theorem 5 holds with weak convergence.

Hilbert spaces satisfy the **Radon Riesz property**:

If u_k converge weakly to u and $\|u_k\| \rightarrow \|u\|$, then $\|u_k - u\| \rightarrow 0$.

So, we have strong convergence as well as weak convergence of solutions.

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

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

Lower semicontinuity: \mathcal{J} is weakly lsc in ℓ^2 .

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

Lower semicontinuity: \mathcal{J} is weakly lsc in ℓ^2 .

Compactness of sublevel sets:

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

Lower semicontinuity: \mathcal{J} is weakly lsc in ℓ^2 .

Compactness of sublevel sets:

- We have $\|\cdot\|_2 \leq \|\cdot\|_1$, so $\mathcal{J}(u) \leq C$ implies $\|u\|_2 \leq C$

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

Lower semicontinuity: \mathcal{J} is weakly lsc in ℓ^2 .

Compactness of sublevel sets:

- We have $\|\cdot\|_2 \leq \|\cdot\|_1$, so $\mathcal{J}(u) \leq C$ implies $\|u\|_2 \leq C$
- bounded sequences have weakly convergent subsequences in ℓ^2 .

So, the sublevel-sets of \mathcal{J} are weakly sequentially compact in ℓ^2 .

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

Lower semicontinuity: \mathcal{J} is weakly lsc in ℓ^2 .

Compactness of sublevel sets:

- We have $\|\cdot\|_2 \leq \|\cdot\|_1$, so $\mathcal{J}(u) \leq C$ implies $\|u\|_2 \leq C$
- bounded sequences have weakly convergent subsequences in ℓ^2 .

So, the sublevel-sets of \mathcal{J} are weakly sequentially compact in ℓ^2 .

Theorem 5 thus guarantees weak convergence in ℓ_2 of solutions.

Example: Bounded variation

Recall $\|u\|_{BV} = \|u\|_{L^1} + TV(u)$. Let $A : L^1(\Omega) \rightarrow L^2(\Omega)$ be continuous and

$$\mathcal{J}(u) = \begin{cases} \|u\|_{BV} & u \in BV(\Omega) \\ +\infty & \text{otherwise} \end{cases}.$$

Example: Bounded variation

Recall $\|u\|_{BV} = \|u\|_{L^1} + TV(u)$. Let $A : L^1(\Omega) \rightarrow L^2(\Omega)$ be continuous and

$$\mathcal{J}(u) = \begin{cases} \|u\|_{BV} & u \in BV(\Omega) \\ +\infty & \text{otherwise} \end{cases}.$$

Lower semi-continuity: TV is lower semi-continuous with respect to L^1 convergence (**exercise**)

Example: Bounded variation

Recall $\|u\|_{BV} = \|u\|_{L^1} + TV(u)$. Let $A : L^1(\Omega) \rightarrow L^2(\Omega)$ be continuous and

$$\mathcal{J}(u) = \begin{cases} \|u\|_{BV} & u \in BV(\Omega) \\ +\infty & \text{otherwise} \end{cases}.$$

Lower semi-continuity: TV is lower semi-continuous with respect to L^1 convergence (**exercise**)

Compactness of sublevel sets:

Theorem 6 (Rellich's compactness theorem)

Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary, and let $(u_n)_n \subset BV(\Omega)$ be such that $\sup_n \|u_n\|_{BV} < \infty$. Then there exists $u \in BV(\Omega)$ and a subsequence $(u_{n_k})_k$ such that $u_{n_k} \rightarrow u$ in $L^1(\Omega)$.

Example: Bounded variation

Recall $\|u\|_{BV} = \|u\|_{L^1} + TV(u)$. Let $A : L^1(\Omega) \rightarrow L^2(\Omega)$ be continuous and

$$\mathcal{J}(u) = \begin{cases} \|u\|_{BV} & u \in BV(\Omega) \\ +\infty & \text{otherwise} \end{cases}.$$

Lower semi-continuity: TV is lower semi-continuous with respect to L^1 convergence (**exercise**)

Compactness of sublevel sets:

Theorem 6 (Rellich's compactness theorem)

Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary, and let $(u_n)_n \subset BV(\Omega)$ be such that $\sup_n \|u_n\|_{BV} < \infty$. Then there exists $u \in BV(\Omega)$ and a subsequence $(u_{n_k})_k$ such that $u_{n_k} \rightarrow u$ in $L^1(\Omega)$.

Therefore, Theorem 5 guarantees strong convergence in L^1 .

Example: Total variation

What if we take $\mathcal{J}(u) = \text{TV}(u)$ on domain Ω ?

Compactness of sublevel sets is problematic as $\mathcal{J}(\alpha\chi_\Omega) = 0$ for all $\alpha \in \mathbb{R}$, but additional compactness can come from the data fidelity term:

What if we take $\mathcal{J}(u) = \text{TV}(u)$ on domain Ω ?

Compactness of sublevel sets is problematic as $\mathcal{J}(\alpha\chi_\Omega) = 0$ for all $\alpha \in \mathbb{R}$, but additional compactness can come from the data fidelity term:

Theorem 3.1 (Poincaré inequality)

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

$$\|u - m(u)\|_{L^p} \leq C \text{TV}(u), \quad \forall u \in BV(\Omega),$$

for all $p \in [1, N/(N-1)]$. This holds for $p = 2$ and $N = 2$.

Example: Total variation

Let $\Omega \subset \mathbb{R}^2$, let $A : L^1(\Omega) \rightarrow L^2(\Omega)$ be a bounded linear operator and suppose that $A\chi_\Omega \neq 0$.

Given u_n s.t. $\text{TV}(u_n) + \frac{1}{2} \|Au_n - f\|_2^2 \leq C$, $|m(u_n)|$ is also uniformly bounded:

- let $w_n = m(u_n)$ and $v_n = u_n - m(u_n)$. Then, $\int v_n = 0$ and $\text{TV}(v_n) = \text{TV}(u_n)$. So, by the Poincaré inequality, $\|v_n\|_{L^2} \leq C'$.
- Observe now that $C \geq \|Au_n - f\|_2 \geq \|Au_n\|_2 - \|f\|_2$, so $\|Au_n\|_2$ is uniformly bounded. Hence

$$C \geq \|Au_n\|_2 = |m(u_n)| \|A\chi_\Omega\|_2 - \|Av_n\|_2.$$

So, Poincaré inequality tells us that $\|u_n\|_{L^2}$ and hence $\|u_n\|_1$ is uniformly bounded, and Rellich's compactness theorem allows us to extract a L^1 convergent subsequence.

Variational regularisation

Background

Regularisation properties

Convergence rates

The dual perspective

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

The subdifferential

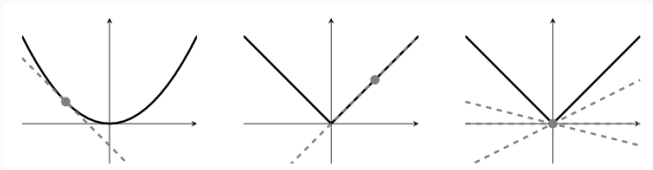
For convex functionals, we can generalise the concept of a derivative for non-differentiable functions.

Definition 7

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

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

is called the subdifferential of E at u .



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

1. If E is differentiable at u , then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$.

1. If E is differentiable at u , then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$. Conversely, if $\partial E(u)$ is a singleton, then E is differentiable at u .

1. If E is differentiable at u , then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$. Conversely, if $\partial E(u)$ is a singleton, then E is differentiable at u .
2. u is a minimiser of E if and only if $0 \in \partial E(u)$.

1. If E is differentiable at u , then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$. Conversely, if $\partial E(u)$ is a singleton, then E is differentiable at u .
2. u is a minimiser of E if and only if $0 \in \partial E(u)$.
3. If $u \in \text{int}(\text{dom}(E))$ and E is a closed convex function, then $\partial E(u) \neq \emptyset$.

1. If E is differentiable at u , then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$. Conversely, if $\partial E(u)$ is a singleton, then E is differentiable at u .
2. u is a minimiser of E if and only if $0 \in \partial E(u)$.
3. If $u \in \text{int}(\text{dom}(E))$ and E is a closed convex function, then $\partial E(u) \neq \emptyset$.
4. If $\text{dom}(E) \neq \emptyset$ and $u \notin \text{dom}(E)$ then $\partial E(u) = \emptyset$.

1. If E is differentiable at u , then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$. Conversely, if $\partial E(u)$ is a singleton, then E is differentiable at u .
2. u is a minimiser of E if and only if $0 \in \partial E(u)$.
3. If $u \in \text{int}(\text{dom}(E))$ and E is a closed convex function, then $\partial E(u) \neq \emptyset$.
4. If $\text{dom}(E) \neq \emptyset$ and $u \notin \text{dom}(E)$ then $\partial E(u) = \emptyset$.
5. If $E : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ is a proper convex function and $u \in \text{dom}(E)$, then $\partial E(u)$ is a weak-* compact convex subset of \mathcal{U}^* .

Calculus rules for the subdifferential

For $\alpha \geq 0$, $\partial(\alpha E)(x) = \alpha \partial E(x)$.

Let $F = E(Ax + b)$ where $A : \mathcal{U} \rightarrow \mathcal{V}$ is a linear operator and $b \in \mathcal{V}$. Then $\partial F(x) = A^* \partial E(Ax + b)$.

Let $E : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ and $F : \mathcal{U} \rightarrow \bar{\mathbb{R}}$ be proper lsc convex functions and suppose that there exists $u \in \text{dom}(E) \cap \text{dom}(F)$ such that E is continuous at u . Then $\partial(E + F) = \partial E + \partial F$.

In general, we have $\partial(E + F) \supset \partial E + \partial F$, but equality may not hold.

As an example, consider

$$E(x) = \begin{cases} -\sqrt{x} & x \geq 0 \\ +\infty & x < 0 \end{cases} \quad \text{and} \quad F(x) = E(-x) = \begin{cases} -\sqrt{-x} & x \leq 0 \\ +\infty & x > 0 \end{cases}$$

Then, $F + E = \iota_{\{0\}}$ and $\partial(F + E)(0) = \mathbb{R}$. On the other hand, $\partial F(0) = \partial E(0) = \emptyset$.

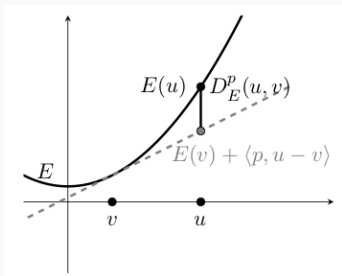
Bregman distances

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 8

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

$$\mathcal{D}_E^p(u, v) = E(u) - E(v) - \langle p, u - v \rangle. \quad (4.1)$$



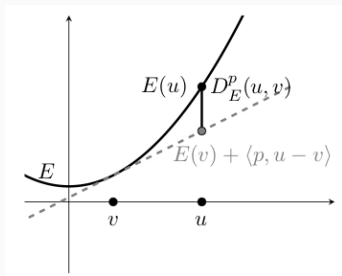
Bregman distances

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 8

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

$$\mathcal{D}_E^p(u, v) = E(u) - E(v) - \langle p, u - v \rangle. \quad (4.1)$$



Example:

For $E(u) = \frac{1}{2} \|u\|^2$, $\partial E(v) = \{v\}$, so

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

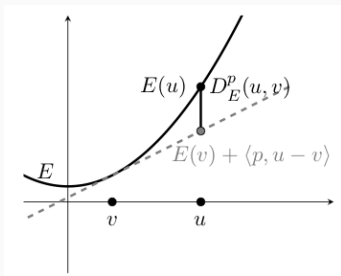
Bregman distances

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 8

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

$$\mathcal{D}_E^p(u, v) = E(u) - E(v) - \langle p, u - v \rangle. \quad (4.1)$$



Example:

For $E(u) = \frac{1}{2} \|u\|^2$, $\partial E(v) = \{v\}$, so

$$\begin{aligned} \mathcal{D}_E^v(u, v) &= \frac{1}{2} \|u\|^2 - \frac{1}{2} \|v\|^2 - \langle v, u - v \rangle \\ &= \frac{1}{2} \|u\|^2 + \frac{1}{2} \|v\|^2 - \langle v, u \rangle \\ &= \frac{1}{2} \|u - v\|^2. \end{aligned}$$

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

Theorem 9

*Assume that the source condition is satisfied at a \mathcal{J} -minimising solution $u_{\mathcal{J}}^{\dagger}$. Let $f = Au_{\mathcal{J}}^{\dagger}$ and let f_{δ} be such that $\|f_{\delta} - f\| \leq \delta$. Let $u_{\delta} \in \operatorname{argmin}_u \Phi_{\alpha, f_{\delta}}(u)$ be a regularised solution. Then, letting $v = A^*p^{\dagger}$, we have*

$$D_{\mathcal{J}}^v(u_{\delta}, u_{\mathcal{J}}^{\dagger}) \leq \frac{1}{2\alpha} \left(\delta + \alpha \|p^{\dagger}\| \right)^2.$$

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial \mathcal{J}(u) = \{u\}$

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial\mathcal{J}(u) = \{u\}$
- the source condition simply says that $u_{\mathcal{J}}^{\dagger} = A^* p^{\dagger}$ for some $p^{\dagger} \in \mathcal{V}$.

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial\mathcal{J}(u) = \{u\}$
- the source condition simply says that $u_{\mathcal{J}}^{\dagger} = A^* p^{\dagger}$ for some $p^{\dagger} \in \mathcal{V}$.
- Recall that $D^{\nu}(u_{\delta}, u_{\mathcal{J}}^{\dagger}) = \frac{1}{2} \left\| u_{\delta} - u_{\mathcal{J}}^{\dagger} \right\|^2$.

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial\mathcal{J}(u) = \{u\}$
- the source condition simply says that $u_{\mathcal{J}}^{\dagger} = A^* p^{\dagger}$ for some $p^{\dagger} \in \mathcal{V}$.
- Recall that $D^{\nu}(u_{\delta}, u_{\mathcal{J}}^{\dagger}) = \frac{1}{2} \left\| u_{\delta} - u_{\mathcal{J}}^{\dagger} \right\|^2$.

Maximum entropy regularisation: On $L^1_+(\Omega)$, consider

$$\mathcal{J}(u) = \int_{\Omega} u \ln u - u dx$$

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial\mathcal{J}(u) = \{u\}$
- the source condition simply says that $u_{\mathcal{J}}^{\dagger} = A^* p^{\dagger}$ for some $p^{\dagger} \in \mathcal{V}$.
- Recall that $D^{\nu}(u_{\delta}, u_{\mathcal{J}}^{\dagger}) = \frac{1}{2} \left\| u_{\delta} - u_{\mathcal{J}}^{\dagger} \right\|^2$.

Maximum entropy regularisation: On $L_+^1(\Omega)$, consider

$$\mathcal{J}(u) = \int_{\Omega} u \ln u - u dx$$

- For u positive, we have $\partial\mathcal{J}(u) = \{\ln u\}$.

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial\mathcal{J}(u) = \{u\}$
- the source condition simply says that $u_{\mathcal{J}}^{\dagger} = A^* p^{\dagger}$ for some $p^{\dagger} \in \mathcal{V}$.
- Recall that $D^{\nu}(u_{\delta}, u_{\mathcal{J}}^{\dagger}) = \frac{1}{2} \left\| u_{\delta} - u_{\mathcal{J}}^{\dagger} \right\|^2$.

Maximum entropy regularisation: On $L^1_+(\Omega)$, consider

$$\mathcal{J}(u) = \int_{\Omega} u \ln u - u dx$$

- For u positive, we have $\partial\mathcal{J}(u) = \{\ln u\}$.
- The source condition says that $u_{\mathcal{J}}^{\dagger} = e^{A^* p^{\dagger}}$ for some $p^{\dagger} \in \mathcal{V}$.

Tikhonov regularisation: Consider $\mathcal{J} = \frac{1}{2} \|u\|^2$.

- $\partial\mathcal{J}(u) = \{u\}$
- the source condition simply says that $u_{\mathcal{J}}^{\dagger} = A^* p^{\dagger}$ for some $p^{\dagger} \in \mathcal{V}$.
- Recall that $D^{\vee}(u_{\delta}, u_{\mathcal{J}}^{\dagger}) = \frac{1}{2} \left\| u_{\delta} - u_{\mathcal{J}}^{\dagger} \right\|^2$.

Maximum entropy regularisation: On $L_+^1(\Omega)$, consider

$$\mathcal{J}(u) = \int_{\Omega} u \ln u - u dx$$

- For u positive, we have $\partial\mathcal{J}(u) = \{\ln u\}$.
- The source condition says that $u_{\mathcal{J}}^{\dagger} = e^{A^* p^{\dagger}}$ for some $p^{\dagger} \in \mathcal{V}$.
- The Bregman distance is the Kullback-Liebert divergence

$$D^{\vee}(u_{\delta}, u_{\mathcal{J}}^{\dagger}) = \int u_{\delta} \ln \left(\frac{u_{\delta}}{u_{\mathcal{J}}^{\dagger}} \right).$$

Sparse regularisation: Consider $\mathcal{J}(u) = \|u\|_1 = \sum_{j \in \mathbb{N}} |u_j|$.

Sparse regularisation: Consider $\mathcal{J}(u) = \|u\|_1 = \sum_{j \in \mathbb{N}} |u_j|$.

$$\blacksquare \partial \mathcal{J}(u) = \left\{ v \in \ell_\infty(\mathbb{N}) ; \|v\|_\infty \leq 1 \quad \text{and} \quad \forall j \in \text{supp}(u), v_j = \text{sign}(u_j) \right\}$$

Sparse regularisation: Consider $\mathcal{J}(u) = \|u\|_1 = \sum_{j \in \mathbb{N}} |u_j|$.

- $\partial \mathcal{J}(u) = \{v \in \ell_\infty(\mathbb{N}) ; \|v\|_\infty \leq 1 \text{ and } \forall j \in \text{supp}(u), v_j = \text{sign}(u_j)\}$
- Suppose that $v \stackrel{\text{def.}}{=} A^* p^\dagger \in \partial \mathcal{J}(u_{\mathcal{J}}^\dagger)$ is such that for some $c \in (0, 1)$,

$$|v_j| < 1 - c, \quad \forall j \notin \text{supp}(u_{\mathcal{J}}^\dagger) \stackrel{\text{def.}}{=} S^\dagger.$$

Sparse regularisation: Consider $\mathcal{J}(u) = \|u\|_1 = \sum_{j \in \mathbb{N}} |u_j|$.

- $\partial \mathcal{J}(u) = \{v \in \ell_\infty(\mathbb{N}) ; \|v\|_\infty \leq 1 \text{ and } \forall j \in \text{supp}(u), v_j = \text{sign}(u_j)\}$
- Suppose that $v \stackrel{\text{def.}}{=} A^* p^\dagger \in \partial \mathcal{J}(u_{\mathcal{J}}^\dagger)$ is such that for some $c \in (0, 1)$,

$$|v_j| < 1 - c, \quad \forall j \notin \text{supp}(u_{\mathcal{J}}^\dagger) \stackrel{\text{def.}}{=} S^\dagger.$$

Then,

$$\begin{aligned} \mathcal{D}_{\mathcal{J}}^v(u_\delta, u_{\mathcal{J}}^\dagger) &= \|u_\delta\|_1 - \|u_{\mathcal{J}}^\dagger\|_1 - \langle v, u_\delta - u_{\mathcal{J}}^\dagger \rangle \\ &= \|u_\delta\|_1 - \langle v, u_\delta \rangle \\ &\geq \|u_\delta\|_1 - \sum_{j \in S^\dagger} |(u_\delta)_j| - (1 - c) \sum_{j \notin S^\dagger} |(u_\delta)_j| = c \sum_{j \notin S^\dagger} |(u_\delta)_j|. \end{aligned}$$

Sparse regularisation: Consider $\mathcal{J}(u) = \|u\|_1 = \sum_{j \in \mathbb{N}} |u_j|$.

- $\partial \mathcal{J}(u) = \{v \in \ell_\infty(\mathbb{N}) ; \|v\|_\infty \leq 1 \text{ and } \forall j \in \text{supp}(u), v_j = \text{sign}(u_j)\}$
- Suppose that $v \stackrel{\text{def.}}{=} A^* p^\dagger \in \partial \mathcal{J}(u_{\mathcal{J}}^\dagger)$ is such that for some $c \in (0, 1)$,

$$|v_j| < 1 - c, \quad \forall j \notin \text{supp}(u_{\mathcal{J}}^\dagger) \stackrel{\text{def.}}{=} S^\dagger.$$

Then,

$$\begin{aligned} \mathcal{D}_{\mathcal{J}}^v(u_\delta, u_{\mathcal{J}}^\dagger) &= \|u_\delta\|_1 - \|u_{\mathcal{J}}^\dagger\|_1 - \langle v, u_\delta - u_{\mathcal{J}}^\dagger \rangle \\ &= \|u_\delta\|_1 - \langle v, u_\delta \rangle \\ &\geq \|u_\delta\|_1 - \sum_{j \in S^\dagger} |(u_\delta)_j| - (1 - c) \sum_{j \notin S^\dagger} |(u_\delta)_j| = c \sum_{j \notin S^\dagger} |(u_\delta)_j|. \end{aligned}$$

This alludes to the fact that ℓ_1 promotes sparse solutions.

Although, as we will see later, in this case, stronger convergence bounds (in terms of $\|\cdot\|_1$) are possible.

1. Since u_δ is a minimiser,

$$\alpha \mathcal{J}(u_\delta) + \frac{1}{2} \|Au_\delta - f_\delta\|^2 \leq \alpha \mathcal{J}(u_\mathcal{J}^\dagger) + \frac{1}{2} \|Au_\mathcal{J}^\dagger - f_\delta\|^2.$$

2. Using the fact that $\|Au_\mathcal{J}^\dagger - f_\delta\| \leq \delta$ and adding and subtracting $\langle A^*p^\dagger, u_\delta - u_\mathcal{J}^\dagger \rangle$ to the LHS of the previous inequality, we obtain

$$\begin{aligned} \alpha D_\mathcal{J}^\vee(u_\delta, u_\mathcal{J}^\dagger) + \frac{1}{2} \|Au_\delta - f_\delta\|^2 + \alpha \langle A^*p^\dagger, u_\delta - u_\mathcal{J}^\dagger \rangle &\leq \frac{\delta^2}{2}. \\ \implies \alpha D_\mathcal{J}^\vee(u_\delta, u_\mathcal{J}^\dagger) + \frac{1}{2} \|Au_\delta - f_\delta\|^2 + \alpha \langle p^\dagger, Au_\delta - Au_\mathcal{J}^\dagger \rangle &\leq \frac{\delta^2}{2} \end{aligned}$$

3. By adding and subtracting $\langle p^\dagger, f_\delta \rangle$, we see that the LHS is precisely

$$\frac{1}{2} \|Au_\delta - f_\delta + \alpha p^\dagger\|^2 + \alpha D_\mathcal{J}^\vee(u_\delta, u_\mathcal{J}^\dagger) - \frac{\alpha^2}{2} \|p^\dagger\|^2 + \alpha \langle p^\dagger, f_\delta - f_\dagger \rangle.$$

4. Rearranging and by Cauchy-Schwarz:

$$D_\mathcal{J}^\vee(u_\delta, u_\mathcal{J}^\dagger) \leq \frac{1}{2\alpha} \left(\delta^2 + \alpha^2 \|p^\dagger\|^2 + 2\alpha \|p^\dagger\| \delta \right).$$

Variational regularisation

Background

Regularisation properties

Convergence rates

The dual perspective

We have so far considered

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

When \mathcal{J} is a convex functional, it is often convenient (both from a theoretical and practical perspective) to consider the **dual** formulation.

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

Definition 10

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

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

- F^* is convex regardless of whether F is convex.
- We have the Fenchel Young inequality: $\langle x, y \rangle \leq F(x) + F^*(y)$,
- if $F : V \subset \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 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 $F^*(y) = \frac{1}{2} \|y\|^2$:

■ for all x ,

$$\begin{aligned}\langle x, y \rangle - \frac{1}{2} \|x\|^2 &\leq \|x\| \|y\| - \frac{1}{2} \|x\|^2 \\ &\leq \frac{1}{2} \|x\|^2 + \frac{1}{2} \|y\|^2 - \frac{1}{2} \|x\|^2 = \frac{1}{2} \|y\|^2.\end{aligned}$$

So, $F^*(y) \leq \frac{1}{2} \|y\|^2$.

■ Setting $x \stackrel{\text{def.}}{=} y$ in the supremum above yields $F^*(y) \geq \frac{1}{2} \|y\|^2$.

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

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

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

Definition 11

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

Examples: norms, the total variation functional.

Properties:

- Clearly $E(0) = 0$.
- 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 $j = 1, \dots, M$.

Consider **the primal problem**:

$$\begin{aligned} \min_{x \in V} F_0(x) \text{ subject to } Ax = y, \\ F_j(x) \leq b_j, \quad j \in [M], \end{aligned} \tag{\mathcal{P}}$$

where

- $F_0 : V \rightarrow (-\infty, +\infty]$ is called the objective function
- $F_j : V \rightarrow (-\infty, +\infty]$ for $j \in [M]$ are called the constraint functions. $A : V \rightarrow Y$ is a continuous linear functional.

The set $K \stackrel{\text{def.}}{=} \{x \in V ; Ax = y, F_j(x) \leq b_j\}$ is called the admissible set.

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

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

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

The **Lagrange dual function** is defined as

$$H(\xi, \nu) \stackrel{\text{def.}}{=} \inf_{x \in V} L(x, \xi, \nu), \quad \xi \in Y^*, \nu \in \mathbb{R}_{\geq 0}^M.$$

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

Properties of the dual function H :

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

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

$$\sup_{\xi \in Y^*, \nu \in \mathbb{R}_{\geq 0}^M} H(\xi, \nu) \text{ subject to } \nu_\ell \geq 0, \ell \in [M]. \quad (\mathcal{D})$$

This optimisation problem is called the **dual problem** and (\mathcal{P}) is called the **primal problem**.

- If D^* is the supremum of (\mathcal{D}) and P^* is the infimum of (\mathcal{P}) , then we have in general $D^* \leq P^*$ (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 $E : Y \rightarrow (-\infty, +\infty]$ and $F : V \rightarrow (-\infty, +\infty]$ are convex functionals, and $A : V \rightarrow Y$ is a continuous linear operator. This is equivalent to

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

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

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

Theorem 12 (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*

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

Theorem 12 (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*

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

Why might we want to look at the dual problem?

- **Computational:** Suppose $A : V \rightarrow Y$ where $V = \mathbb{R}^N$ and $Y = \mathbb{R}^m$. If A is some measurement operator, we often have $m \ll N$. Now, the primal problem optimises over $V = \mathbb{R}^N$ while the dual problem optimises over $Y = Y^* = \mathbb{R}^m$.
- **Theoretical** insights (see later).

Primal and dual formulations

Back to our problem:

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

Let $E(Au) = \frac{1}{2} \|Au - f_\delta\|^2$ 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 $K = \partial \mathcal{J}(0)$, and $(\alpha \mathcal{J})^*(v) = \alpha \mathcal{J}^*(\alpha^{-1}v)$.

Therefore, the dual problem is

$$\sup_v \langle v, f_\delta \rangle - \frac{1}{2} \|v\|^2 + \iota_K \left(\frac{A^*v}{\alpha} \right) \quad (\mathcal{D}_\alpha)$$

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

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

We have $\langle v, f_\delta \rangle - \frac{1}{2} \|v\|^2 = -\frac{1}{2} \|v - f_\delta\|^2 + \frac{1}{2} \|f_\delta\|^2$, so the dual solution is the projection of f_δ onto a closed convex set, and is therefore **unique**.

The limit primal and dual problems

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

$$\inf_{u: Au=f} \mathcal{J}(u) = \inf_{u \in \mathcal{U}} \iota_{\{f\}}(Au) + \mathcal{J}(u) \quad (\mathcal{P}_0)$$

and

$$\sup_{v: A^* v \in \partial \mathcal{J}(0)} \langle f, v \rangle = - \inf_v \langle -f, v \rangle + \iota_{\partial J(0)}(A^* v) \quad (\mathcal{D}_0)$$

We cannot directly apply Theorem 12 to (\mathcal{P}_0) to deduce strong duality, because $\iota_{\{f\}}$ is not continuous at f .

However, if $\mathcal{J} : \mathcal{U} \rightarrow [0, \infty]$ absolute one-homogeneous and coercive, we can show that (\mathcal{P}_0) is the dual of (\mathcal{D}_0) . So, studying the two problems are still equivalent. **See the additional exercises.**

Theorem 13

Suppose that the source condition holds at a \mathcal{J} -minimising solution $u_{\mathcal{J}}^{\dagger}$. Then, p_{α} the solution to (\mathcal{D}_{α}) with data f is uniformly bounded in α . Moreover, $p_{\alpha} \rightarrow p^{\dagger}$ strongly in \mathcal{V} as $\alpha \rightarrow 0$, where p^{\dagger} is a solution to (\mathcal{D}_0) with smallest norm.

The source condition implies dual convergence

Theorem 13

Suppose that the source condition holds at a \mathcal{J} -minimising solution $u_{\mathcal{J}}^{\dagger}$. Then, p_{α} the solution to (\mathcal{D}_{α}) with data f is uniformly bounded in α . Moreover, $p_{\alpha} \rightarrow p^{\dagger}$ strongly in \mathcal{V} as $\alpha \rightarrow 0$, where p^{\dagger} is a solution to (\mathcal{D}_0) with smallest norm.

Remark: Let u_{α} solve (\mathcal{D}_{α}) with data f and u_{δ} solve (\mathcal{D}_{α}) with f_{δ} .

NB: $u_{\delta} = P_K(f_{\delta}/\alpha)$ the projection onto $K \stackrel{\text{def.}}{=} \{p ; A^*p \in \partial\mathcal{J}(0)\}$.

So,

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

and p_{δ} converges to p^{\dagger} as $\delta/\alpha \rightarrow 0$ and $\alpha \rightarrow 0$.

The minimal norm certificate and structural stability

The dual solutions $p_{\alpha,\delta}$ converge to the minimal norm dual solution p^\dagger as $\alpha, \delta \rightarrow 0$ (with $\delta/\alpha \leq c$). This often means that $A^* p^\dagger$ control the structural properties of $u_{\alpha,\delta}$ for small α and δ .

Example Let $\mathcal{J} = \|\cdot\|_1$ in \mathbb{R}^n . Suppose that $A^* p^\dagger \in \partial J(u^\dagger)$ satisfies $\|(A^* p^\dagger)_{S^c}\|_\infty < 1$ for $S \stackrel{\text{def.}}{=} \text{Supp}(u^\dagger)$.

Claim: $\text{Supp}(u_{\alpha,\delta}) = S$ for α and δ sufficiently small:

- $A^* p \in \partial J(u)$ means that $\|A^* p\|_\infty \leq 1$ and $(A^* p)_S = \text{sign}(u_S)$.
- If $\|(A^* p^\dagger)_{S^c}\|_\infty < 1$, then $\|(A^* p_{\alpha,\delta})_{S^c}\|_\infty < 1$ for all α, δ sufficiently small. This means $\text{Supp}(u_{\alpha,\delta}) \subseteq S$.
- Since we have convergence of $u_{\alpha,\delta}$ to u , we actually have $\text{Supp}(u_{\alpha,\delta}) = S$.

Similar notions of structural stability (stability of level curves) for $\mathcal{J} = TV$.

Proof of Theorem 13 (dual convergence)

Step 1, show that p_α is uniformly bounded in α : Since p_α solves (\mathcal{D}_α) , we have

$$\langle f, p_\alpha \rangle - \frac{\alpha}{2} \|p_\alpha\|^2 \geq \langle f, p^\dagger \rangle - \frac{\alpha}{2} \|p^\dagger\|^2. \quad (5.1)$$

Moreover, as p^\dagger is a solution to (\mathcal{D}_0) , we have $\langle f, p^\dagger \rangle \geq \langle f, p_\alpha \rangle$. So, $\|p^\dagger\| \geq \|p_\alpha\|$.

Step 2, extract a convergent subsequence to some point p_* : 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 $\alpha \rightarrow 0$ in (5.1) yields $\langle f, p_* \rangle \geq \langle y, p^\dagger \rangle$.

Proof of Theorem 13 (dual convergence)

Step 3, show p_* is a solution to \mathcal{D}_0 : Note that $A^*p_{\alpha_{n_k}}$ converges weak-* to A^*p_* , and so $A^*p_* \in \partial\mathcal{J}(0)$ (since this is a weak-* closed set). So, p_* is a solution to (\mathcal{D}_0) .

Step 4, show p_* is of minimal norm: By lower semicontinuity of the norm,

$$\|p_*\| \leq \liminf_k \|p_{\alpha_{n_k}}\| \leq \|p^\dagger\|,$$

and hence, $p_* = p^\dagger$. Moreover, since $\|p_{\alpha_{n_k}}\| \rightarrow \|p^\dagger\|$, by the Radon Riesz property, $p_{\alpha_{n_k}} \rightarrow p_0$ strongly in \mathcal{H} .

Step 5, the entire sequence converges: We have $\lim_{\delta \rightarrow 0} \|p_\alpha - p^\dagger\| = 0$, since otherwise, we can extract a subsequence p_{α_k} such that $\|p_{\alpha_k} - p^\dagger\| > \varepsilon$ and by the above argument, extract a further subsequence which converges strongly to p^\dagger .

We studied variational regularisers of the form

$$R_\alpha(f) = \operatorname{argmin}_u \alpha \mathcal{J}(u) + \frac{1}{2} \|Au - f\|^2.$$

which is a natural generalisation of Tikhonov regularisation.

- This is a convergent regularisation under appropriate continuity properties of A , \mathcal{J} is proper, lsc with compact sublevel sets and $\delta^2/\alpha(\delta) \rightarrow 0$.
- We introduced a source condition for studying convergence rates:
 - this gives convergence rates in terms of Bregman distances under a source condition.
 - For convex regularisers, we saw how to reformulate using the dual problem. The source condition is simply saying that the limit dual problem ($\alpha \rightarrow 0$) has a solution.
 - The source condition guarantees dual convergence, and this can provide finer notions of convergence.