Inverse Problems
Variational regularisation

Clarice Poon
University of Bath

March 1, 2020

Outline

Variational regularisation

Background

Regularisation properties

Convergence rates

The dual perspective

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

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

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

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

where

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

Example: smoothing regularisers

Let $\mathcal{J}(u) = \|Lu\|_{\mathcal{Z}}$ where $L : \mathcal{U} \to \mathcal{Z}$ 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 $L : \mathcal{U} \to \mathcal{Z}$ 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 $L : \mathcal{U} \to \mathcal{Z}$ 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

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

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 $L : \mathcal{U} \to \mathcal{Z}$ 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 $\mathcal{J}(u) = \begin{cases} \|u\|_1 & u \in \ell_1(\mathbb{N}) \\ +\infty & u \in \ell_2(\mathbb{N}) \setminus \ell_2(\mathbb{N}) \end{cases}$.

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 $W : \ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})$. For example, W is some wavelet transform.

Example: Lasso

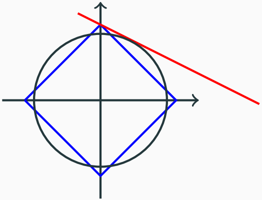
Consider $\mathcal{U} = \mathcal{V} = \ell_2(\mathbb{N})$ and $\mathcal{J}(u) = \begin{cases} \|u\|_1 & u \in \ell_1(\mathbb{N}) \\ +\infty & u \in \ell_2(\mathbb{N}) \setminus \ell_2(\mathbb{N}) \end{cases}$.

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 $W : \ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})$. 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

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

Let $\|u\|_{BV} \overset{\text{def.}}{=} \|u\|_{L^1} + \mathrm{TV}(u)$, and the space of bounded variations $\left\{ u \in L^1 \; ; \; \mathrm{TV}(u) < \infty \right\}$ 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, $TV(\chi_C) = \mathrm{Per}(C)$.

Bayesian viewpoint of variational methods

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.

Bayesian viewpoint of variational methods

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

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

Bayesian viewpoint of variational methods

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 $\prod_{i,j} \frac{u_{i,j}^{f_{i,j}}}{f_{i,j}!} e^{-u_{i,j}}$ 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 $R_\alpha(f) \to u^\dagger$?
- Do we have convergent regularisers?
- Convergence rates?

Outline

Infinity

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

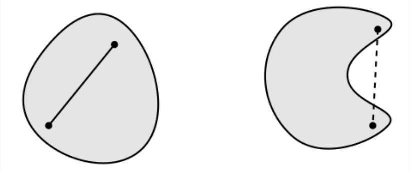
- Useful to model constraints. E.g. if $E : [-1, \infty) \to \mathbb{R}^2$ maps $x \mapsto x^2$, consider instead $\bar{E} : \mathbb{R} \to \bar{\mathbb{R}}$ defined by $\bar{E}(x) = E(x)$ for $x \in [-1, \infty)$ and $\bar{E}(x) = +\infty$ otherwise. No need to worry if $E(x + y)$ is well-defined.
-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

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

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

Convexity

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.

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 $\langle p,\, u\rangle \overset{\text{def.}}{=} p(u)$. The dual space is a Banach space equipped with the norm

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

Banach spaces are complete, normed vector spaces.

Bi-dual

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

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

$E : \mathcal{U} \to \mathcal{U}^{**}$ 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 $\mathcal{U}$.

Dual spaces

Banach spaces are complete, normed vector spaces.

Adjoint

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

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

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 exists $u \in \ell^2$ such that $\|e_j - u\| \to 0$.

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 $u \in \mathcal{U}$.

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 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.

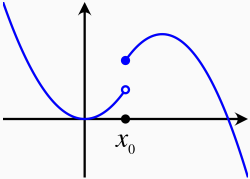
Lower semi-continuity

One useful property is the notion of sequential lower semicontinuity:

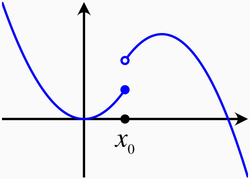
Let $\mathcal{X}$ be a Banach space with topology $\tau_{\mathcal{X}}$. The functional $E : \mathcal{X} \to [-\infty, \infty]$ is said to be sequentially lower semi-continuous with respect to $\tau_{\mathcal{X}}$ at $u \in \mathcal{X}$ if

$$E(x) \leqslant \liminf_{j \to \infty} E(x_j)$$

for all sequences $\{x_j\}_j \subset \mathcal{X}$ with $x_j \to x$ 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 $u^j \to u$ 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,

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

Example: Lower semi-continuity

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

Example: Lower semi-continuity

The functional $\|\cdot\|_1 : \ell^2 \to [0, \infty]$ 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 $\|\cdot\|_1 : \ell^2 \to [0, \infty]$ 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 $u^j \to u$ in ℓ_2, we have

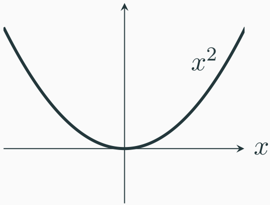
$$u_k^j = \langle e_k,\, u^j \rangle \to \langle e_k,\, u \rangle = u_k.$$

So, by Fatou's lemma

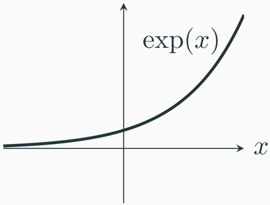
$$\|u\|_1 = \sum_k \lim_{j\to\infty} \left| u_k^j \right| \leqslant \liminf_{j\to\infty} \sum_k \left| u_k^j \right| = \liminf_{j\to\infty} \left\| u^j \right\|_1 .$$

Minimising functionals

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.



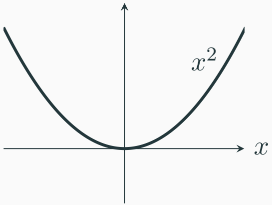
Coercive



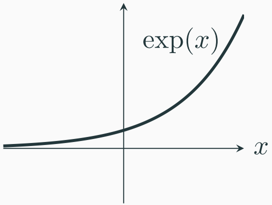
Not coercive.

Minimising functionals

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.



Coercive



Not coercive.

Coercivity is sufficient to ensure boundedness of minimising sequences:

Lemma 2.1

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.

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 $E : \mathcal{U} \to \bar{\mathbb{R}}$ 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 $E : \mathcal{U} \to \bar{\mathbb{R}}$ 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.

Minimising functionals

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 $E : \mathcal{U} \to \bar{\mathbb{R}}$ 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_{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.

$\square$

Minimising functionals

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

Outline

Well-posedness and regularisation properties

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 $A : \mathcal{U} \to \mathcal{V}$ 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 $R_\alpha : \mathcal{V} \to \mathcal{U}$ 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?

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) *$A : \mathcal{U} \to \mathcal{V}$ 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') *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.*

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) $A : \mathcal{U} \to \mathcal{V}$ *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') *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.

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

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

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

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

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!).

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 $f_j \to f$ 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, $\mathcal{J}(u_j) \to \mathcal{J}(u)$.

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 $f_j \to f$ 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, $\mathcal{J}(u_j) \to \mathcal{J}(u)$.

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

(b) From (a), we have
$\Phi_{\alpha,f}(u) \leqslant \|Au - g\|^2 + \|g - f\|^2 + 2\alpha\mathcal{J}(u) \leqslant 2\Phi_{\alpha,g}(u) + \|f-g\|^2.$

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

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

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

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

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 $\Phi_{\alpha,f_j}(u_j) \to \Phi_{\alpha,f}(u)$. So,

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

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

Let

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

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

Let

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

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

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

Let

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

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

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

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

Let

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

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

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

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

Let

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

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

- As $\mathcal{V}$ is a Hilbert space, $\mathbb{L}_f \stackrel{\text{def.}}{=} \left\{v \,;\, v \in \operatorname{argmin}_{u\in\mathcal{U}} \|Au - f\|_{\mathcal{V}}\right\}$ 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.

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) $A : \mathcal{U} \to \mathcal{V}$ *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, there exists a $\mathcal{J}$-minimising solution $u^{\dagger}_{\mathcal{J}}$.

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) *$A : \mathcal{U} \to \mathcal{V}$ 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, there exists a $\mathcal{J}$-minimising solution $u^\dagger_\mathcal{J}$.

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

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

Theorem 5

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

Theorem 5

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

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 a $\mathcal{J}$-minimising solution.

Theorem 5

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

- Since u_δ is a minimiser: $\|Au_\delta - f_\delta\|^2 + \alpha(\delta)\mathcal{J}(u_\delta) \leqslant \frac{1}{2}\left\|Au_\mathcal{J}^\dagger - f_\delta\right\|^2 + \alpha(\delta)\mathcal{J}(u_\mathcal{J}^\dagger)$. This implies that $\mathcal{J}(u_\delta) \leqslant \mathcal{J}(u_\mathcal{J}^\dagger) + \frac{\delta^2}{2\alpha(\delta)}$.
- by compactness of the sublevel sets of $\mathcal{J}$, up to a subsequence u_{δ_n} converges to u_* as $\delta_n \to 0$. By continuity of A, $Au_{\delta_n} \overset{\tau_\mathcal{V}}{\to} Au_*$.
- $u_* \in \mathbb{L}_f$ follows by lsc of $\|\|_\mathcal{V}$ wrt $\tau_\mathcal{V}$ and by minimality of u_{δ_n}:

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

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

Examples: Tikhonov

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

Examples: Tikhonov

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

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

Examples: Tikhonov

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.

Examples: Tikhonov

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.

Examples: Tikhonov

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 $\|u_k\| \to \|u\|$, then $\|u_k - u\| \to 0$.

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

Examples: Lasso

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.

Examples: Lasso

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 $\|\cdot\|_2 \leqslant \|\cdot\|_1$, so $\mathcal{J}(u) \leqslant C$ implies $\|u\|_2 \leqslant C$

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 $\|\cdot\|_2 \leqslant \|\cdot\|_1$, so $\mathcal{J}(u) \leqslant C$ implies $\|u\|_2 \leqslant C$
- 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 $\|\cdot\|_2 \leqslant \|\cdot\|_1$, so $\mathcal{J}(u) \leqslant C$ implies $\|u\|_2 \leqslant C$
- 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.

Recall $\|u\|_{BV} = \|u\|_{L^1} + TV(u)$. Let $A : L^1(\Omega) \to L^2(\Omega)$ 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 $A : L^1(\Omega) \to L^2(\Omega)$ 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 $A : L^1(\Omega) \to L^2(\Omega)$ 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 $u_{n_k} \to u$ in $L^1(\Omega)$.

Example: Bounded variation

Recall $\|u\|_{BV} = \|u\|_{L^1} + TV(u)$. Let $A : L^1(\Omega) \to L^2(\Omega)$ 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 $u_{n_k} \to u$ in $L^1(\Omega)$.

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

Example: Total variation

What if we take $\mathcal{J}(u) = \mathrm{TV}(u)$ 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:

Example: Total variation

What if we take $\mathcal{J}(u) = \mathrm{TV}(u)$ 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 $m(u) = \frac{1}{|\Omega|}\int_\Omega u(x)\mathrm{d}x$. 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 $p \in [1, N/(N-1)]$. This holds for $p = 2$ and $N = 2$.

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

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

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

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

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.

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.

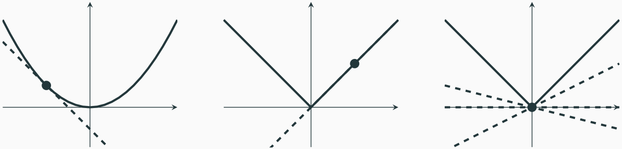
For convex functionals, we can generalise the concept of a derivative for non-differentiable functions.

Definition 7

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

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

is called the subdifferential of E at u.



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

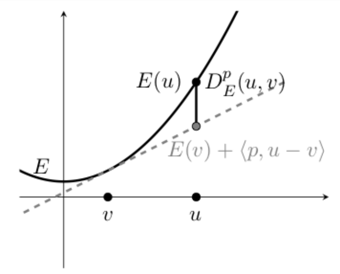
- If E is differentiable at u, then $\partial E(u) \stackrel{\text{def.}}{=} \{\nabla E(u)\}$.
- Let $E : \mathcal{U} \to \bar{\mathbb{R}}$ and $F : \mathcal{U} \to \bar{\mathbb{R}}$ be proper lsc convex functions and suppose that there exists $u \in \mathrm{dom}(E) \cap \mathrm{dom}(F)$ such that E 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 $\mathcal{U}^*$.

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 generalised Bregman distance is given by

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

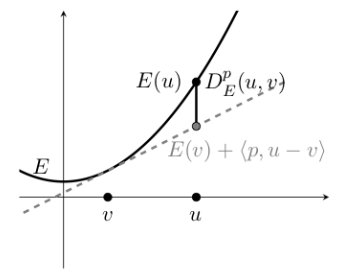


Convergence rates are typically derived via analysis of the *dual problem* and studied in terms of the *Bregman distances* associated with the (convex) regularisation functional $\mathcal{J}$.

Definition 8

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

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



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

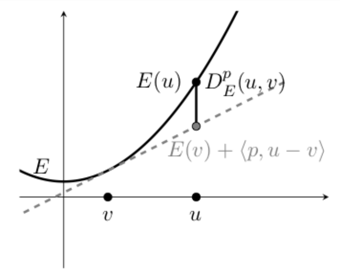
$$\mathcal{D}_E^v(u, v) = \frac{1}{2}\|u\|^2 - \frac{1}{2}\|v\|^2 - \langle v,\, u - v \rangle$$

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 generalised Bregman distance is given by

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



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

Theorem 9

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

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

Remarks on Theorem 9

- For $\mathcal{J} = \frac{1}{2}\|u\|^2$, $\partial\mathcal{J}(u) = \{u\}$ and the source condition simply says that $u_{\mathcal{J}}^\dagger = A^* w$ for some w. Recall that $D^v(u_\delta, u_{\mathcal{J}}^\dagger) = \frac{1}{2}\left\|u_\delta - u_{\mathcal{J}}^\dagger\right\|^2$.
- Another popular regulariser is maximum entropy regularisation,

$$\mathcal{J}(u) = \int_\Omega u \ln u - u \mathrm{d}x$$

defined on $L_+^1(\Omega)$. For u positive, we have $\partial\mathcal{J}(u) = \{\ln u\}$. So the source condition says that $u_{\mathcal{J}}^\dagger = e^{A^* w}$ for some $w \in \mathcal{V}$. The Bregman distance is the Kullback-Lieber divergence

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

Remarks on Theorem 9

Consider $\mathcal{J}(u) = \|u\|_1 = \sum_{j\in\mathbb{N}} |u_j|$. Then, since

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

Suppose that $v \stackrel{\text{def.}}{=} A^* p^\dagger \in \partial\mathcal{J}(u^\dagger_{\mathcal{J}})$ is such that for some $c \in (0,1)$, $|v_j| < 1 - c$ for all $j \notin \operatorname{supp}(u^\dagger_{\mathcal{J}}) \stackrel{\text{def.}}{=} S^\dagger$. Then,

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

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.

Proof of Theorem 9

1. Since u_δ is a minimiser,
   $\alpha\mathcal{J}(u_\delta) + \frac{1}{2}\|Au_\delta - f_\delta\|^2 \leqslant \alpha\mathcal{J}(u^\dagger_{\mathcal{J}}) + \frac{1}{2}\left\|Au^\dagger_{\mathcal{J}} - f_\delta\right\|^2.$
2. Using the fact that $\left\|Au^\dagger_{\mathcal{J}} - f_\delta\right\| \leqslant \delta$ and adding and subtracting $\langle A^* p^\dagger,\, u_\delta - u^\dagger_{\mathcal{J}}\rangle$ to the LHS of the previous inequality, we obtain
$$\alpha D^v_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) + \frac{1}{2}\|Au_\delta - f_\delta\|^2 + \alpha\langle A^* p^\dagger,\, u_\delta - u^\dagger_{\mathcal{J}}\rangle \leqslant \frac{\delta^2}{2}.$$
$$\implies \alpha D^v_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) + \frac{1}{2}\|Au_\delta - f_\delta\|^2 + \alpha\langle p^\dagger,\, Au_\delta - Au^\dagger_{\mathcal{J}}\rangle \leqslant \frac{\delta^2}{2}$$
3. By adding and subtracting $\langle p^\dagger,\, f_\delta\rangle$, we see that the LHS is precisely
$$\frac{1}{2}\left\|Au_\delta - f_\delta + \alpha p^\dagger\right\|^2 + \alpha D^v_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) - \frac{\alpha^2}{2}\left\|p^\dagger\right\|^2 + \alpha\langle p^\dagger,\, f_\delta - f_\dagger\rangle.$$
4. Rearranging and by Cauchy-Schwarz:
$$D^v_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) \leqslant \frac{1}{2\alpha}\left(\delta^2 + \alpha^2\left\|p^\dagger\right\|^2 + 2\alpha\left\|p^\dagger\right\|\delta\right).$$

Outline

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 $F : V \to (-\infty, +\infty]$, its convex conjugate is $F^* : V^* \to (-\infty, +\infty]$ 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: $\langle x, y\rangle \leqslant F(x) + F^*(y)$,
- if $F : V \subset \mathbb{R}^n \to (-\infty, +\infty]$ is convex and lower semi-continuous, then $F^{**} = F$.
- if F is convex, then $y \in \partial F(x)$ if and only if $F(x) + F^*(y) = \langle x, y\rangle$.

(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,

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

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

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

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 $E(0) = 0$.

Examples: $\|\cdot\|_p$, 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 $E(u) = \langle p, u \rangle$.

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**:

$$\min_{x \in V} F_0(x) \text{ subject to } Ax = y, \tag{5.1}$$

$$F_j(x) \leqslant b_j,\ j \in [M], \tag{5.2}$$

where

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

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

Primal and dual formulations

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

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

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 $H(\xi,\nu) = -\infty$.

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 $H(\xi, \nu) \leqslant \inf_{x \in K} F_0(x)$ 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

$$\sup_{\xi \in Y^*, \nu \in \mathbb{R}^M} H(\xi, \nu) \text{ subject to } \nu_\ell \geqslant 0,\ \ell \in [M]. \tag{5.3}$$

This optimisation problem is called the **dual problem** and (5.1) is called the **primal problem**.

- If D^* is the supremum of (5.3) and P^* is the infimum of (5.1), 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

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

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

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

Theorem 11 (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*

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

Primal and dual formulations

Theorem 11 (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 $A : V \to Y$ 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 $V = RR^N$ while the dual problem optimises over $Y = Y^* = \mathbb{R}^m$.
- **Theoretical** insights (see later).

Primal and dual formulations

Back to our problem:

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

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

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

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 $\delta \to 0$ are

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

and

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

We cannot directly apply Theorem 11 to $(\mathcal{P}_0)$ to deduce strong duality, because $\iota_{\{f\}}$ is not continuous at f.

However, if $\mathcal{J} : \mathcal{U} \to [0, \infty]$ 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.

The source condition implies dual convergence

Theorem 12

Suppose that the source condition holds at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$. Then, p_α the solution to $(\mathcal{D}_\alpha)$ 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 $(\mathcal{D}_0)$ with smallest norm.

The source condition implies dual convergence

Theorem 12

Suppose that the source condition holds at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$. Then, p_α the solution to $(\mathcal{D}_\alpha)$ 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 $(\mathcal{D}_0)$ with smallest norm.

Remark: Let u_α solve $(\mathcal{D}_\alpha)$ with data f and u_δ solve$(\mathcal{D}_\alpha)$ with f_δ.

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

So,

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

and p_δ converges to $p^\dagger$ as $\delta/\alpha \to 0$ and $\alpha \to 0$.

The minimal norm certificate and structural stability

The dual solutions $p_{\alpha,\delta}$ converge to the minimal norm dual solution $p^\dagger$ as $\alpha, \delta \to 0$ (with $\delta/\alpha \leqslant c$). 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 $\left\|(A^* p^\dagger)_{S^c}\right\|_\infty < 1$ for $S \overset{\text{def.}}{=} \mathrm{Supp}(u^\dagger)$.

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

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

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

Proof of Theorem 12 (dual convergence)

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

$$\langle f,\, p_\alpha\rangle - \frac{\alpha}{2}\left\|p_\alpha\right\|^2 \geqslant \langle f,\, p^\dagger\rangle - \frac{\alpha}{2}\left\|p^\dagger\right\|^2. \tag{5.4}$$

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

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 $\alpha \to 0$ in (5.4) yields $\langle f,\, p_*\rangle \geqslant \langle y,\, p^\dagger\rangle$.

Proof of Theorem 12 (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,

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

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

Step 5, the entire sequence converges: We have $\lim_{\delta \to 0} \left\| p_\alpha - p^\dagger \right\| = 0$, 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 $p^\dagger$.

Summary

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 $\delta^2/\alpha(\delta) \to 0$.
- 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 ($\alpha \to 0$) has a solution.
  - The source condition guarantees dual convergence, and this can provide finer notions of convergence.