Part III: Inverse Problems

Lecture notes, Michaelmas 2024
University of Cambridge

Ferdia Sherry
ferdiasherry.com

November 11, 2024

Contents

These lecture notes are for Part III "Inverse Problems", taught by Ferdia Sherry in Michaelmas 2024 at the University of Cambridge. These notes are based on the notes for Part III "Inverse Problems" taught by Yury Korolev and Jonas Latz in Michaelmas 2020 at the University of Cambridge [1]. Complementary material can be found in the following books, lecture notes and review papers:

1. Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Springer, 1996.

2. Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier and Frank Lenzen. *Variational Methods in Imaging*. Springer, 2008.

3. Kristian Bredies and Dirk Lorenz. *Mathematical Image Processing*. Springer, 2018

4. Martin Benning and Martin Burger. *Modern regularization methods for inverse problems*. Acta Numerica, 2018.

5. Bryan P. Rynne and Martin A. Youngson, *Linear Functional Analysis*, Springer, 2008.

6. Masoumeh Dashti and Andrew M. Stuart, *The Bayesian approach to inverse problems*, Handbook of Uncertainty Quantification, 2016.

7. Jari Kaipio and Erkki Somersalo, *Statistical and computational inverse problems*, vol. 160 of Applied Mathematical Sciences, 2005.

8. O. Kallenberg, *Foundations of modern probability theory*, Springer, 1997.

9. Andrew M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numerica, 2010.

These lecture notes are under constant redevelopment and might contain typos or errors. I would very much appreciate if you report any mistakes found to fs436@cam.ac.uk. Thanks!

[1] https://www.damtp.cam.ac.uk/research/cia/inverse-problems-michaelmas-2020

Chapter 1

Introduction to Inverse Problems

Inverse problems arise from the need to gain information about an unknown object of interest from given indirect measurements. Inverse problems have several applications varying from medical imaging and industrial process monitoring to ozone layer tomography and modelling of financial markets. The common feature for inverse problems is the need to understand indirect measurements and to overcome extreme sensitivity to noise and modelling inaccuracies. In this course we employ both deterministic and probabilistic approach to inverse problems to find stable and meaningful solutions that allow us quantify how inaccuracies in the data or model affect the obtained estimate.

1.1 Well-posed and ill-posed problems

We start by considering the problem of finding $u \in \mathbf{R}^d$ that satisfies the equation

$$f = Au, \tag{1.1}$$

where $f \in \mathbf{R}^k$ is given. We refer to f as observed data or measurement and u as an unknown. The physical phenomena that relates the unknown and the measurement is modelled by a matrix $A \in \mathbf{R}^{k\times d}$. In real life, the perfect data given in (1.1) is perturbed by noise and we observe measurements

$$f_\eta = Au + \eta, \tag{1.2}$$

where $\eta \in \mathbf{R}^k$ represents the observational noise.

We are interested in ill-posed inverse problems, where the inverse problem is more difficult to solve than the direct problem of finding f_η when u is given. To explain this, we first need to introduce well-posedness as defined by Jacques Hadamard [14]:

Definition 1.1. *A problem is called well-posed if*

1. *There exists at least one solution. (Existence)*
2. *There is at most one solution. (Uniqueness)*
3. *The solution depends continuously on data. (Stability)*

The direct or forward problem is assumed to be well-posed. The inverse problems are ill-posed and break at least one of the above conditions.

1. Assume that $d < k$ and $A : \mathbf{R}^d \to \mathcal{R}(A) \subsetneq \mathbf{R}^k$, where the range of A is a proper subset of $\mathbf{R}^k$. Furthermore, we assume that A has a unique inverse $A^{-1} : \mathcal{R}(A) \to \mathbf{R}^k$. Because of the noise in the measurement $f_\eta \notin \mathcal{R}(A)$ so that simply inverting A with the data given in (1.2) is not possible. Note that usually only the statistical properties of the noise n are known so we cannot just subtract it.

2. Assume next that $d > k$ and $A : \mathbf{R}^d \to \mathbf{R}^k$, in which case the system is underdetermined. We then have more unknowns than equations which means that there are several possible solutions.

3. Consider next the case $d = k$, in which there exists $A^{-1} : \mathbf{R}^k \to \mathbf{R}^k$. The condition number $\kappa = \lambda_1/\lambda_k$, where λ_1 and λ_k are the biggest and smallest eigenvalues of A, may be very large. Such a matrix is said to be ill-conditioned and is almost singular. In this case the problem is sensitive even to smallest errors in the measurement. Hence the naïve reconstruction $\widetilde{u} = A^{-1}f_\eta = u + A^{-1}\eta$ does not produce a meaningful solution but will be dominated by $A^{-1}\eta$. Note that $\|A^{-1}\eta\|_2 \approx \|\eta\|_2/\lambda_k$ can be arbitrarily large.

The last part illustrates one of the key questions of inverse problem theory: how can we stabilise the reconstruction process while maintaining acceptable accuracy?

A deterministic way of achieving a unique and stable solution for the problem (1.2) is to use regularisation theory. In the classical Tikhonov regularisation a solution is attained by solving

$$\min_{u\in\mathbf{R}^d} \left(\left\| Au - f_\eta \right\|^2 + \alpha \|Lu\|^2 \right). \tag{1.3}$$

Above, α acts as a tuning parameter balancing the effect of the data fidelity term $\left\|Au - f_\eta\right\|^2$ and the stabilising regularisation term $\|u\|^2$. The first half of the course will concentrate on regularisation theory.

Another way of tackling problems arising from ill-posedness is Bayesian inversion. The idea of statistical inversion methods is to rephrase the inverse problem as a question of statistical inference. We then consider the problem

$$f_\eta = Au + \eta, \tag{1.4}$$

where the measurement, unknown and noise are now modelled as random variables. This approach allows us to model the noise through its statistical properties. We can also encode our *a priori* knowledge of the unknown in form of a probability distribution that assigns higher probability to those values of u we expect to see. The solution to (1.4) is so-called *posterior distribution*, which is the conditional probability distribution of u given a measurement f_η. This distribution can then be used to obtain estimates that are most likely in some sense. We will return to the Bayesian approach to inverse problems in the second half of the course

In this course we will concentrate on continuous inverse problems where in (1.1) and (1.2) $A : \mathcal{X} \to \mathcal{Y}$ is a linear or non-linear forward operator acting between some spaces $\mathcal{X}$ and $\mathcal{Y}$, typically Hilbert or Banach spaces, the measured data $f_\eta \in \mathcal{Y}$ is a function and $u \in \mathcal{X}$ is the quantity we want to reconstruct from the data. Linear inverse problems include such important applications as computer tomography, magnetic resonance imaging and image deblurring in microscopy or astronomy. In other important applications, such as seismic imaging, the forward operator is non-linear (e.g., parameter identification problems for PDEs). Next we will take a look at some examples of linear and non-linear inverse problems to see what kind of challenges we face when trying to solve them.

1.2 Examples of inverse problems

1.2.1 Signal deblurring

The deblurring (or deconvolution) problem of recovering an input signal u from an observed signal

$$f_\eta(t) = \int_{-\infty}^{\infty} a(t-s)u(s)\,\mathrm{d}s + \eta(t)$$

occurs in many imaging, and image- and signal processing applications. Here the function a is known as the blurring kernel.

The noiseless data is given by $f(t) = \int_{-\infty}^{\infty} a(t-s)u(s)\,\mathrm{d}s$ and its Fourier transform is $\widehat{f}(\xi) = \int_{-\infty}^{\infty} \exp(-i\xi t)f(t)dt$. The convolution theorem implies

$$\widehat{f}(\xi) = \widehat{a}(\xi)\widehat{u}(\xi),$$

and hence by inverse Fourier transform

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(it\xi)\frac{\widehat{f}(\xi)}{\widehat{a}(\xi)}\,\mathrm{d}\xi.$$

However, we can only observe noisy measurements and hence we have, in the frequency domain, $\widehat{f_\eta}(\xi) = \widehat{a}(\xi)\widehat{u}(\xi) + \widehat{\eta}(\xi)$. The estimate u_{est} based on the convolution theorem is given by

$$u_{\text{est}}(t) = u(t) + \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(it\xi)\frac{\widehat{\eta}(\xi)}{\widehat{a}(\xi)}\,\mathrm{d}\xi,$$

which is often not even well defined, since usually the kernel a decreases exponentially (or has compact support), making the denominator small, whereas the Fourier transform of the noise will be non-zero.

1.2.2 Heat equation

Next, we study the problem of recovering the initial condition u of the heat equation from a noisy observation f_n of the solution at some time $T > 0$. We consider the heat equation on a torus $\mathbf{T}^d = (\mathbf{R}/\mathbf{Z})^d$, with Dirichlet boundary conditions

$$\begin{cases} \frac{\partial v}{\partial t} - \Delta v = 0 & \text{on } \mathbf{T}^d \times \mathbf{R}_+ \\ v(x,t) = 0 & \text{on } \partial\mathbf{T}^d \times \mathbf{R}_+ \\ v(x,T) = f(x) & \text{on } \mathbf{T}^d \\ v(x,0) = u(x) & \text{on } \mathbf{T}^d \end{cases}$$

where Δ denotes the Laplace operator and $\mathcal{D}(\Delta) = H_0^1(\mathbf{T}^d) \bigcap H^2(\mathbf{T}^d)$. Note that the operator $-\Delta$ is positive and self-adjoint on Hilbert space $\mathcal{H} = L^2(\mathbf{T}^d)$.

Given a function $u \in L^2(\mathbf{T}^d)$ we can decompose it as a Fourier series

$$u(x) = \sum_{n\in\mathbf{Z}^d} \widehat{u_n}\exp(2\pi i\langle n, x\rangle),$$

where $\widehat{u_n} = \langle u, \exp(2\pi i\langle n, \cdot\rangle)\rangle_{L^2(\mathbf{T}^d)}$ are the Fourier coefficients, and the identity holds for almost every $x \in \mathbf{T}^d$. The L^2-norm of u is given by the Parseval's identity $\|u\|_{L^2}^2 = \sum |u_n|^2$. Remember that

the Sobolev space $H^s(\mathbf{T}^d)$, $s \in \mathbf{N}$, consists of all $L^2(\mathbf{T}^d)$ integrable functions whose α^{th} order weak derivatives exist and are $L^2(\mathbf{T}^d)$ integrable for all $|\alpha| \leqslant s$. The fractional Sobolev space $H^s(\mathbf{T}^d)$ is given by the subspace of functions $u \in L^2(\mathbf{T}^d)$, such that

$$\|u\|_{H^s}^2 = \sum_{n \in \mathbf{Z}^d} (1 + 4\pi^2|n|^2)^s |u_n|^2 < \infty. \tag{1.5}$$

Note that for a positive integer s, the above definition agrees with the definition given using the weak derivatives. For $s < 0$, we define $H^s(\mathbf{T}^d)$ via duality or as the closure of $L^2(\mathbf{T}^d)$ under the norm (1.5). The resulting spaces are separable for all $s \in \mathbf{R}$.

The eigenvectors of $-\Delta$ in $\mathbf{T}^d$ form the orthonormal basis of $L^2(\mathbf{T}^d)$ and the eigenvalues are given by $4\pi^2|n|^2$, $n \in \mathbf{Z}^d$. We can also work on real-valued functions where the eigenfunctions $\{\varphi_j\}_{j=1}^{\infty}$ comprise sine and cosine functions. The eigenvalues of $-\Delta$, when ordered as a sequence, then satisfy $\lambda_j \asymp j^{2/d}$. The notation $\asymp$ means that there exist constants $C_1, C_2 > 0$, such that $C_1 j^{2/d} \leqslant \lambda_j \leqslant C_2 j^{2/d}$.

The solution to the forward heat equation can be written as

$$v(t) = \sum_{j=1}^{\infty} u_j \exp(-\lambda_j t)\varphi_j.$$

We notice that

$$\|v(t)\|_{H^s}^2 \asymp \sum_{j=1}^{\infty} j^{\frac{2s}{d}} \exp(-2\lambda_j t)|u_j|^2 \asymp t^{-s} \sum_{j=1}^{\infty} (\lambda_j t)^s \exp(-2\lambda_j t)|u_j|^2 \leqslant Ct^{-s} \sum_{j=1}^{\infty} |u_j|^2 = Ct^{-s}\|u\|_{L^2}$$

which implies that $v(t) \in H^s(\mathbf{T}^d)$ for all $s > 0$.

We now have the observation model

$$f_\eta = Au + \eta,$$

where $A = \exp(T\Delta)$ and η is the observational noise. The noise is not usually smooth (the often assumed white noise is not even an L^2 function), and hence the measurement f_η is not in the image space $\operatorname{im}(\exp(T\Delta)) \subset \cap_{s>0} H^s(\mathbf{T}^d)$.

1.2.3 Differentiation

Consider the problem of evaluating the derivative of a function $f \in L^2[0, \pi/2]$. Let

$$Df = f',$$

where $D\colon L^2[0, \pi/2] \to L^2[0, \pi/2]$.

Proposition 1.2. *The operator D is unbounded from $L^2[0, \pi/2] \to L^2[0, \pi/2]$.*

Proof. Take a sequence $f_n(x) = \sin(nx)$, $n = 1, \ldots, \infty$. Clearly, $f_n \in L^2[0, \pi/2]$ for all n and $\|f_n\| = \sqrt{\frac{\pi}{4}}$. However, $Df_n(x) = n\cos(nx)$ and $\|Df_n\| = n \to \infty$ as $n \to \infty$. Therefore, D is unbounded. □

This shows that differentiation is ill posed when considered as an operator from L^2 to L^2. It does not mean that it can not be well-posed in other spaces. For instance, it is well-posed from H^1 (the Sobolev space of L^2 functions whose derivatives are also in L^2) to L^2. Indeed, $\forall u \in H^1$ we get

$$\|Df\|_{L^2} = \|f'\|_{L^2} \leqslant \|f\|_{H^1} \asymp \|f\|_{L^2} + \|f'\|_{L^2}.$$

However, since in practice we typically deal with functions corrupted by non-smooth noise, the setting of L^2 is relevant to practice, while the H^1 setting is not. Differentiation can be written as an inverse problem for an integral equation. For instance, the derivative u of some function $f \in L^2[0,1]$ with $f(0) = 0$ satisfies

$$f(x) = \int_0^x u(t)\,\mathrm{d}t,$$

which can be written as an operator equation $Au = f$ with $(A\cdot)(x) := \int_0^x \cdot(t)\,\mathrm{d}t$.

1.2.4 Matrix inversion

In finite dimensions, the inverse problem (1.1) is a linear system. Linear systems are formally well posed, in the sense that the error in the solution is bounded by some constant times the error in the right-hand side. However, this constant depends on the condition number of the matrix A and can get arbitrary large for matrices with large condition numbers. In this case, we speak of *ill-conditioned* problems.

Consider the problem (1.1) with $u \in \mathbf{R}^n$ and $f \in \mathbf{R}^n$ being n-dimensional vectors with real entries and $A \in \mathbf{R}^{n\times n}$ being a matrix with real entries. Assume further that A is symmetric and positive definite. We know from the spectral theory of symmetric matrices that there exist eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0$ and corresponding (orthonormal) eigenvectors $a_j \in \mathbf{R}^n$ for $j \in \{1,\ldots,n\}$ such that A can be written as

$$A = \sum_{j=1}^{n} \lambda_j a_j a_j^\top. \tag{1.6}$$

It is well known from numerical linear algebra that the condition number $\kappa = \lambda_1/\lambda_n$ is a measure of how stably (1.1) can be solved, which we will illustrate in what follows.

We assume that we measure f_δ rather than f, with $\|f - f_\delta\|_2 \leqslant \delta\|A\| = \delta\lambda_1$, where $\|\cdot\|_2$ denotes the Euclidean norm of $\mathbf{R}^n$ and $\|A\|$ the operator norm of A (which equals the largest eigenvalue of A). Then, if we further denote with u_δ the solution of $Au_\delta = f_\delta$, the difference between u_δ and the solution u to (1.1) is

$$u - u_\delta = \sum_{j=1}^{n} \lambda_j^{-1} a_j a_j^\top (f - f_\delta).$$

Therefore, we can estimate

$$\|u - u_\delta\|_2^2 = \sum_{j=1}^{n} \lambda_j^{-2} \underbrace{\|a_j\|_2^2}_{=1} |a_j^\top (f - f_\delta)|^2 \leqslant \lambda_n^{-2} \|f - f_\delta\|_2^2,$$

due to the orthonormality of eigenvectors, the Cauchy–Schwarz inequality, and $\lambda_n \leqslant \lambda_j$. Thus, taking square roots on both sides yields the estimate

$$\|u - u_\delta\|_2 \leqslant \lambda_n^{-1} \|f - f_\delta\|_2 \leqslant \kappa\delta.$$

Hence, we observe that in the worst case an error δ in the data y is amplified by the condition number κ of the matrix A. A matrix with large κ is therefore called *ill-conditioned*. Let us demonstrate the effect of this error amplification with a small example.

Example 1.1. Consider the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & \frac{1001}{1000} \end{pmatrix},$$

which has eigenvalues $\lambda_j = 1 + \frac{1}{2000} \pm \sqrt{1 + \frac{1}{2000^2}}$, condition number $\kappa \approx 4002 \gg 1$, and operator norm $\|A\| \approx 2$. For given data $f = (1,1)^\top$ the solution to $Au = f$ is $u = (1,0)^\top$. Now let us instead consider perturbed data $f_\delta = (99/100, 101/100)^\top$. The solution u_δ to $Au_\delta = f_\delta$ is then $u_\delta = (-19.01, 20)^\top$. Let us reflect on the amplification of the measurement error. By our initial assumption we find that $\delta = \|f - f_\delta\|/\|A\| \approx \|(0.01, -0.01)^\top\|/2 = \sqrt{2}/200$. Moreover, the norm of the error in the reconstruction is then $\|u - u_\delta\| = \|(20.01, 20)^\top\| \approx 20\sqrt{2}$. As a result, the amplification due to the perturbation is $\|u - u_\delta\|/\delta \approx 4000 \approx \kappa$.

1.2.5 Tomography

In almost any tomography application, the underlying inverse problem is either the inversion of the Radon transform[1] or of the X-ray transform. For $u \in C_0^\infty(\mathbf{R}^n)$, $s \in \mathbf{R}$, and $\theta \in S^{n-1}$ the *Radon transform* $R : C_0^\infty(\mathbf{R}^n) \to C^\infty(S^{n-1} \times \mathbf{R})$ can be defined as the integral operator

$$\begin{aligned} f(\theta, s) = (Ru)(\theta, s) &= \int_{x \cdot \theta = s} u(x)\, \mathrm{d}x \\ &= \int_{\theta^\perp} u(s\theta + y)\, \mathrm{d}y, \end{aligned} \tag{1.7}$$

which, for $n = 2$, coincides with the X-ray transform,

$$f(\theta, s) = (Pu)(\theta, s) = \int_{\mathbf{R}} u(s\theta + t\theta^\perp)\, \mathrm{d}t,$$

for $\theta \in S^{n-1}$ and $\theta^\perp$ being the vector orthogonal to θ. Hence, the X-ray transform (and therefore also the Radon transform in two dimensions) integrates the function u over lines in $\mathbf{R}^n$, see Fig. 1.1.

Example 1.2. Let $n = 2$. Then S^{n-1} is simply the unit sphere $S^1 = \{\theta \in \mathbf{R}^2 \mid \|\theta\| = 1\}$. We can choose for instance $\theta = (\cos(\varphi), \sin(\varphi))^\top$, for $\varphi \in [0, 2\pi)$, and parametrise the Radon transform in terms of φ and s, i.e.

$$f(\varphi, s) = (Ru)(\varphi, s) = \int_{\mathbf{R}} u(s\cos(\varphi) - t\sin(\varphi), s\sin(\varphi) + t\cos(\varphi))\, \mathrm{d}t. \tag{1.8}$$

Note that—with respect to the origin of the reference coordinate system—φ determines the angle of the line along one wants to integrate, while s is the offset from that line from the centre of the coordinate system. It can be shown that the Radon transform is linear and continuous, i.e. $R \in \mathcal{L}(L^2(B), L^2(Z))$, and even compact.

In **X-ray Computed Tomography (CT)**, the unknown quantity u represents a spatially varying density that is exposed to X-radiation from different angles, and that absorbs the radiation according to its material or biological properties. The basic modelling assumption for the intensity decay of an X-ray beam is that within a small distance Δt it is proportional to the intensity itself, the density, and the distance, i.e.

$$\frac{I(x + (t + \Delta t)\theta) - I(x + t\theta)}{\Delta t} = -I(x + t\theta)u(x + t\theta) + O(\Delta t),$$

[1] Named after the Austrian mathematician Johann Karl August Radon (16 December 1887 – 25 May 1956).

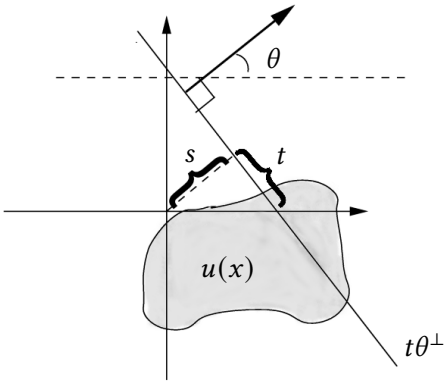

Figure 1.1: Visualisation of the Radon transform in two dimensions[2] (which coincides with the X-ray transform). The function u is integrated over the ray parametrised by θ and s.

for $x \in \theta^\perp$. By taking the limit $\Delta t \to 0$ we end up with the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} I(x + t\theta) = -I(x + t\theta)u(x + t\theta), \tag{1.9}$$

Let $r > 0$ be the radius of the domain of interest centred at the origin. Then, we integrate (1.9) from $t = -\sqrt{r^2 - \|x\|_2^2}$, the position of the emitter, to $t = \sqrt{r^2 - \|x\|_2^2}$, the position of the detector, and obtain

$$\int_{-\sqrt{r^2-\|x\|_2^2}}^{\sqrt{r^2-\|x\|_2^2}} \frac{\frac{\mathrm{d}}{\mathrm{d}t} I(x + t\theta)}{I(x + t\theta)} \,\mathrm{d}t = -\int_{-\sqrt{r^2-\|x\|_2^2}}^{\sqrt{r^2-\|x\|_2^2}} u(x + t\theta) \,\mathrm{d}t.$$

Note that, since $\mathrm{d}/\mathrm{d}x \log(f(x)) = f'(x)/f(x)$, the left hand side in the above equation simplifies to

$$\int_{-\sqrt{r^2-\|x\|_2^2}}^{\sqrt{r^2-\|x\|_2^2}} \frac{\frac{\mathrm{d}}{\mathrm{d}t} I(x + t\theta)}{I(x + t\theta)} \,\mathrm{d}t = \log\left(I\left(x + \sqrt{r^2 - \|x\|_2^2}\theta\right)\right) - \log\left(I\left(x - \sqrt{r^2 - \|x\|_2^2}\theta\right)\right).$$

As we know the radiation intensity at both the emitter and the detector, we therefore know $f(x, \theta) = \log(I(x - \theta\sqrt{r^2 - \|x\|_2^2})) - \log(I(x + \theta\sqrt{r^2 - \|x\|_2^2}))$ and we can write the estimation of the unknown density u as the inverse problem of the X-ray transform (1.8) (if we further assume that u can be continuously extended to zero outside of the circle of radius r).

1.2.6 Groundwater flow/hydraulic tomography

One goal in hydraulic tomography is to estimate the permeability of a groundwater reservoir. The permeability describes the conductivity of the groundwater reservoir and is, e.g., used to estimate the travel time of toxic or radioactive particles in the groundwater. To estimate the permeability, the water pressure in several positions within the reservoir is measured. Pressure head and permeability are linked through Darcy's law and the (assumed) incompressibility of water.

[2] Figure adapted from Wikipedia `https://commons.wikimedia.org/w/index.php?curid=3001440`, by Begemotv2718, CC BY-SA 3.0.

Let $D \subseteq \mathbf{R}^d$ $(d = 1, 2, 3)$ be an open, bounded, connected set with smooth boundary representing the groundwater reservoir. Let $a : \overline{D} \to (0, \infty)$ be a continuously differentiable function representing the permeability and let $s : \overline{D} \to \mathbf{R}$ be a continuous function representing the water sources in the reservoir. Furthermore, assume that the water pressure is 0 outside of D. Darcy's law states that the pressure $p : D \to \mathbf{R}$, the flux $\vec{q} : D \to \mathbf{R}^d$, and the permeability in the reservoir are related as follows:

$$\vec{q}(x) = -a(x)\nabla p(x) \qquad (x \in D).$$

Incompressibility on the other hand requires that the divergence of the flux is fully controlled by in- and outflow given through the source term s:

$$\nabla \cdot \vec{q}(x) = s(x) \qquad (x \in D).$$

Finally, we can combine these assertions and obtain the elliptic partial differential equation

$$\begin{aligned} -\nabla \cdot (a(x)\nabla p(x)) &= s(x) && (x \in D) \\ p(x) &= 0 && (x \in \partial D). \end{aligned}$$

In the described set-up, we now observe the pressure p in several positions $x_1, \ldots, x_I \in D$, e.g., we observe $f_\eta = (p(x_i) : i = 1, \ldots, I) + \eta$. We consider the inverse problem consisting in the estimation of the permeability a using the pressure measurements f_η. Indeed, using noisy point evaluations of the solution of the partial differential equation, we try to estimate its diffusion coefficient. Note that the map $a \mapsto (p(x_i) : i = 1, \ldots, I)$ is non-linear. Hence, this inverse problem is a non-linear inverse problem.

Chapter 2

Generalised Solutions

Functional analysis is the basis of the theory that we will cover in this course. We cannot recall all basic concepts of functional analysis and instead refer to popular textbooks that deal with this subject, e.g., [7, 22, 19]. Nevertheless, we will recall a few important definitions that will be used in this lecture.

We will focus on inverse problems with *bounded linear operators* A, i.e. $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ with

$$\|A\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} := \sup_{u \in \mathcal{X}\setminus\{0\}} \frac{\|Au\|_{\mathcal{Y}}}{\|u\|_{\mathcal{X}}} = \sup_{\|u\|_{\mathcal{X}} \leqslant 1} \|Au\|_{\mathcal{Y}} < \infty.$$

For $A\colon \mathcal{X} \to \mathcal{Y}$ we further denote by

1. $\operatorname{dom}(A) := \mathcal{X}$ the domain of A,
2. $\ker(A) := \{u \in \mathcal{X} \mid Au = 0\}$ the kernel of A,
3. $\operatorname{im}(A) := \{f \in \mathcal{Y} \mid \exists u \in \mathcal{X},\ f = Au\}$ the range of A.

We say that A is continuous at $u \in \mathcal{X}$ if for all $\varepsilon > 0$ there exists $\delta > 0$ with

$$\|Au - Av\|_{\mathcal{Y}} \leqslant \varepsilon \text{ for all } v \in \mathcal{X} \text{ with } \|u - v\|_{\mathcal{X}} \leqslant \delta.$$

For linear K it can be shown that continuity is equivalent to boundedness, i.e. the existence of a constant $C > 0$ such that

$$\|Au\|_{\mathcal{Y}} \leqslant C\|u\|_{\mathcal{X}}$$

for all $u \in \mathcal{X}$. Note that the optimal constant C actually equals the operator norm $\|A\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})}$.

In this Chapter we only consider $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ with $\mathcal{X}$ and $\mathcal{Y}$ being Hilbert spaces. Every Hilbert space $\mathcal{U}$ is equipped with an *inner product*, which we are going to denote by $\langle \cdot, \cdot \rangle_{\mathcal{U}}$ (or simply $\langle \cdot, \cdot \rangle$, whenever the space is clear from the context). In analogy to the transpose of a matrix, this inner product structure together with the theorem of Fréchet-Riesz [22, Section 2.10, Theorem 2.E] allows us to define the *adjoint operator* of A, denoted with $A^* \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$, as the unique solution to the following identity:

$$\langle Au, v \rangle_{\mathcal{Y}} = \langle u, A^* v \rangle_{\mathcal{X}}, \text{ for all } u \in \mathcal{X}, v \in \mathcal{Y}.$$

In addition to this, the inner product is used to define orthogonality. Two elements $u, v \in \mathcal{X}$ are said to be *orthogonal* if $\langle u, v \rangle = 0$. For a subset $\mathcal{X}' \subset \mathcal{X}$ the *orthogonal complement* of $\mathcal{X}'$ in $\mathcal{X}$ is defined as

$$\mathcal{X}'^{\perp} := \{u \in \mathcal{X} \mid \langle u, v \rangle_{\mathcal{X}} = 0 \text{ for all } v \in \mathcal{X}'\}.$$

One can show that $\mathcal{X}'^{\perp}$ is a closed subspace and that $\mathcal{X}^{\perp} = \{0\}$. Moreover, we have that $\mathcal{X}' \subset (\mathcal{X}'^{\perp})^{\perp}$. If $\mathcal{X}'$ is a closed subspace then we even have $\mathcal{X}' = (\mathcal{X}'^{\perp})^{\perp}$. In this case, we can give an *orthogonal decomposition* of $\mathcal{X}$:

$$\mathcal{X} = \mathcal{X}' \oplus \mathcal{X}'^{\perp}.$$

By this notation, we mean that every element $u \in \mathcal{X}$ can uniquely be represented as

$$u = x + x^{\perp} \text{ with } x \in \mathcal{X}' \text{ and } x^{\perp} \in \mathcal{X}'^{\perp},$$

see for instance [22, Section 2.9, Corollary 1]. The mapping $u \mapsto x$ defines a linear operator $P_{\mathcal{X}'} \in \mathcal{L}(\mathcal{X}, \mathcal{X})$, which is called the *orthogonal projection* on $\mathcal{X}'$.

Lemma 2.1 (cf. [15, Section 5.16])**.** *Let $\mathcal{X}' \subset \mathcal{X}$ be a closed subspace. The orthogonal projection onto $\mathcal{X}'$, $P_{\mathcal{X}'}$ satisfies the following conditions:*

1. *$P_{\mathcal{X}'}$ is self-adjoint, i.e. $P^*_{\mathcal{X}'} = P_{\mathcal{X}'}$,*
2. *$\|P_{\mathcal{X}'}\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} = 1$ if $\mathcal{X}' \neq \{0\}$,*
3. *$I - P_{\mathcal{X}'} = P_{\mathcal{X}'^{\perp}}$,*
4. *$\|u - P_{\mathcal{X}'}u\|_{\mathcal{X}} \leqslant \|u - v\|_{\mathcal{X}}$ for all $v \in \mathcal{X}'$,*
5. *$x = P_{\mathcal{X}'}u$ if and only if $x \in \mathcal{X}'$ and $u - x \in \mathcal{X}'^{\perp}$.*

Remark 2.2. Note that for a non-closed subspace $\mathcal{X}'$ we only have $(\mathcal{X}'^{\perp})^{\perp} = \overline{\mathcal{X}'}$. For $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ we therefore have

- $\operatorname{im}(A)^{\perp} = \ker(A^*)$ and thus $\ker(A^*)^{\perp} = \overline{\operatorname{im}(A)}$,
- $\operatorname{im}(A^*)^{\perp} = \ker(A)$ and thus $\ker(A)^{\perp} = \overline{\operatorname{im}(A^*)}$.

Hence, we can deduce the following orthogonal decompositions

$$\mathcal{X} = \ker(A) \oplus \overline{\operatorname{im}(A^*)} \text{ and } \mathcal{Y} = \ker(A^*) \oplus \overline{\operatorname{im}(A)}.$$

2.1 Generalised Inverses

Recall the inverse problem

$$Au = f, \tag{2.1}$$

where $A\colon \mathcal{X} \to \mathcal{Y}$ is a linear bounded operator and $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.

Definition 2.3 (Minimal-norm solutions)**.** *An element $u \in \mathcal{X}$ is called*

- *a least-squares solution of* (2.1) *if*

$$\|Au - f\|_{\mathcal{Y}} = \inf\{\|Av - f\|_{\mathcal{Y}} | v \in \mathcal{X}\},$$

- *a minimum-norm solution of* (2.1) *(and is denoted by $u^{\dagger}$) if it is a least-squares solution, and*

$$\|u^{\dagger}\|_{\mathcal{X}} \leqslant \|v\|_{\mathcal{X}} \quad \text{for all least-squares solutions } v.$$

Remark 2.4. Since $\mathrm{im}(A)$ is not closed in general (it is never closed for a compact operator, unless the range is finite-dimensional), a least-squares solution may not exist. If it exists, then the minimum-norm solution is unique (it is the orthogonal projection of the zero element onto the non-empty closed convex set defined by $\|Au - f\|_{\mathcal{Y}} = \min\{\|Av - f\|_{\mathcal{Y}} | v \in \mathcal{X}\}$).

In numerical linear algebra it is a well known fact that the normal equation can be used to compute least-squares solutions. The same holds true in the infinite-dimensional case.

Theorem 2.5. *Let $f \in \mathcal{Y}$ and $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then, the following three assertions are equivalent for any $u \in \mathcal{X}$.*

1. $Au = P_{\overline{\mathrm{im}(A)}} f$.
2. *u is a least-squares solution of the inverse problem* (2.1).
3. *u solves the* normal equation
$$A^* Au = A^* f. \tag{2.2}$$

Remark 2.6. The name normal equation is derived from the fact that for any solution u its residual $Au - f$ is orthogonal (normal) to $\mathrm{im}(A)$. This can be readily seen, as we have for any $v \in \mathcal{X}$ that

$$0 = \langle v, A^*(Au - f)\rangle_{\mathcal{X}} = \langle Av, Au - f\rangle_{\mathcal{Y}}$$

which shows $Au - f \in \mathrm{im}(A)^\perp$.

Proof of Theorem 2.5. For $1 \Rightarrow 2$: Let $u \in \mathcal{X}$ such that $Au = P_{\overline{\mathrm{im}(A)}} f$ and let $v \in \mathcal{X}$ be arbitrary. With the basic properties of the orthogonal projection, Lemma 2.1, point 4, we have

$$\|Au - f\|_{\mathcal{Y}} = \|P_{\overline{\mathrm{im}(A)}} f - f\|_{\mathcal{Y}} \leqslant \inf_{g \in \overline{\mathrm{im}(A)}} \|g - f\|_{\mathcal{Y}} \leqslant \inf_{g \in \mathrm{im}(A)} \|g - f\|_{\mathcal{Y}} = \inf_{v \in \mathcal{X}} \|Av - f\|_{\mathcal{Y}},$$

which shows that u is a least-squares solution.

For $2 \Rightarrow 3$: Let $u \in \mathcal{X}$ be a least-squares solution and let $v \in \mathcal{X}$ an arbitrary element. We define the quadratic polynomial $F\colon \mathbf{R} \to \mathbf{R}$,

$$F(\lambda) := \|A(u + \lambda v) - f\|_{\mathcal{Y}}^2 = \lambda^2 \|Av\|_{\mathcal{Y}}^2 - 2\lambda \langle Av, f - Au\rangle_{\mathcal{Y}} + \|f - Au\|_{\mathcal{Y}}^2.$$

A necessary condition for $u \in \mathcal{X}$ to be a least-squares solution is $F'(0) = 0$, which leads to $\langle v, A^*(f - Au)\rangle_{\mathcal{X}} = 0$. As v was arbitrary, it follows that the normal equation (2.2) must hold.

For $3 \Rightarrow 1$: From the normal equation it follows that $A^*(f - Au) = 0$, which is equivalent to $f - Au \in \mathrm{im}(A)^\perp$, see Remark 2.6. Since $\mathrm{im}(A)^\perp = \left(\overline{\mathrm{im}(A)}\right)^\perp$ and $Au \in \mathrm{im}(A) \subset \overline{\mathrm{im}(A)}$, the assertion follows from Lemma 2.1, point 5:

$$Au = P_{\overline{\mathrm{im}(A)}} f \Leftrightarrow Au \in \overline{\mathrm{im}(A)} \text{ and } f - Au \in \left(\overline{\mathrm{im}(A)}\right)^\perp.$$

□

Lemma 2.7. *Let $f \in \mathcal{Y}$ and let $\mathbf{L}$ be the set of least-squares solutions to the inverse problem* (2.1). *Then, $\mathbf{L}$ is non-empty if and only if $f \in \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$.*

Proof. Let $u \in \mathbf{L}$. It is easy to see that $f = Au + (f - Au) \in \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$ as the normal equation are equivalent to $f - Au \in \mathrm{im}(A)^\perp$.

Consider now $f \in \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$. Then there exists $u \in \mathcal{X}$ and $g \in \mathrm{im}(A)^\perp = \left(\overline{\mathrm{im}(A)}\right)^\perp$ such that $f = Au + g$ and thus $P_{\overline{\mathrm{im}(A)}} f = P_{\overline{\mathrm{im}(A)}} Au + P_{\overline{\mathrm{im}(A)}} g = Au$ and the assertion follows from Theorem 2.5, point 1. □

Remark 2.8. If $\mathrm{im}(A)$ is finite-dimensional, then $\mathrm{im}(A)$ is closed, i.e. $\overline{\mathrm{im}(A)} = \mathrm{im}(A)$. Thus, when the measurements are finite-dimensional, there always exists a least-squares solution.

Theorem 2.9. *Let $f \in \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$. Then there exists a unique minimum-norm solution $u^\dagger$ to the inverse problem* (2.1) *and all least-squares solutions are given by $\{u^\dagger\} + \ker(A)$.*

Proof. From Lemma 2.7 we know that there exists a least-squares solution. As noted in Remark 2.4, in this case the minimum-norm solution is unique. Let φ be an arbitrary least-squares solution. Using Theorem 2.5 we get

$$A(\varphi - u^\dagger) = A\varphi - Au^\dagger = P_{\overline{\mathrm{im}(A)}} f - P_{\overline{\mathrm{im}(A)}} f = 0, \tag{2.3}$$

which shows that $\varphi - u^\dagger \in \ker(A)$, hence the assertion. □

If a least-squares solution exists for a given $f \in \mathcal{Y}$, then the minimum-norm solution can be computed (at least in theory) using the Moore–Penrose generalised inverse.

Definition 2.10. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ and let*

$$\widetilde{A} \coloneqq A|_{\ker(A)^\perp} : \ker(A)^\perp \to \mathrm{im}(A)$$

denote the restriction of A to $\ker(A)^\perp$. The Moore–Penrose inverse *$A^\dagger$ is defined as the unique linear extension of $\widetilde{A}^{-1}$ to*

$$\mathrm{dom}(A^\dagger) = \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$$

with

$$\ker(A^\dagger) = \mathrm{im}(A)^\perp.$$

Remark 2.11. Due to the restriction to $\ker(A)^\perp$ and $\mathrm{im}(A)$ we have that $\widetilde{A}$ is injective and surjective. Hence, $\widetilde{A}^{-1}$ exists and is linear and — as a consequence — $A^\dagger$ is well-defined on $\mathrm{im}(A)$. Moreover, due to the orthogonal decomposition $\mathrm{dom}(A^\dagger) = \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$, there exist for arbitrary $f \in \mathrm{dom}(A^\dagger)$ elements $f_1 \in \mathrm{im}(A)$ and $f_2 \in \mathrm{im}(A)^\perp$ with $f = f_1 + f_2$. Therefore, we have

$$A^\dagger f = A^\dagger f_1 + A^\dagger f_2 = A^\dagger f_1 = \widetilde{A}^{-1} f_1 = \widetilde{A}^{-1} P_{\overline{\mathrm{im}(A)}} f, \tag{2.4}$$

where we have used that $f_2 \in \mathrm{im}(A)^\perp = \ker(A^\dagger)$. Thus, $A^\dagger$ is well-defined on the entire domain $\mathrm{dom}(A^\dagger)$.

Remark 2.12. As orthogonal complements are always closed we get that

$$\overline{\mathrm{dom}(A^\dagger)} = \overline{\mathrm{im}(A)} \oplus \mathrm{im}(A)^\perp = \mathcal{Y},$$

and hence, $\mathrm{dom}(A^\dagger)$ is dense in $\mathcal{Y}$. Thus, if $\mathrm{im}(A)$ is closed it follows that $\mathrm{dom}(A^\dagger) = \mathcal{Y}$ and on the other hand, $\mathrm{dom}(A^\dagger) = \mathcal{Y}$ implies $\mathrm{im}(A)$ is closed. We note that for ill-posed problems $\mathrm{im}(A)$ is usually not closed; for instance, if A is compact then $\mathrm{im}(A)$ is closed if and only if it is finite-dimensional [1, Ex.1 Section 7.1].

If A is bijective, we have that $A^\dagger = A^{-1}$. We also highlight that the extension $A^\dagger$ is not necessarily continuous.

Example 2.13. To illustrate the definition of the Moore–Penrose inverse we consider a simple example in finite dimensions. Let the linear operator $A\colon \mathbf{R}^3 \to \mathbf{R}^2$ be given by

$$Ax = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix}.$$

It is easy to see that $\operatorname{im}(A) = \{f \in \mathbf{R}^2 \mid f_2 = 0\}$ and $\ker(A) = \{x \in \mathbf{R}^3 \mid x_1 = 0\}$. Thus, $\ker(A)^\perp = \{x \in \mathbf{R}^3 \mid x_2, x_3 = 0\}$. Therefore, $\widetilde{A}\colon \ker(A)^\perp \to \operatorname{im}(A)$, given by $x \mapsto (2x_1, 0)^\top$, is bijective and its inverse $\widetilde{A}^{-1}\colon \operatorname{im}(A) \to \ker(A)^\perp$ is given by $f \mapsto (f_1/2, 0, 0)^\top$.

To get the Moore–Penrose inverse $A^\dagger$, we need to extend $\widetilde{A}^{-1}$ to $\operatorname{im}(A) \oplus \operatorname{im}(A)^\perp$ in such a way that $A^\dagger f = 0$ for all $f \in \operatorname{im}(A)^\perp = \{f \in \mathbf{R}^2 \mid f_1 = 0\}$. It is easy to see that the Moore–Penrose inverse $A^\dagger\colon \mathbf{R}^2 \to \mathbf{R}^3$ is given by the following expression

$$A^\dagger f = \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} f_1/2 \\ 0 \\ 0 \end{pmatrix}.$$

Let us consider data $\widetilde{f} = (8, 1)^\top \notin \operatorname{im}(A)$. Then, $A^\dagger \widetilde{f} = A^\dagger (8, 1)^\top = (4, 0, 0)^\top$.

Let us show that $A^\dagger$ can be characterised by the Moore–Penrose identities:

Theorem 2.14 ([12, Proposition 2.3])**.** *The Moore–Penrose inverse $A^\dagger$ satisfies $\operatorname{im}(A^\dagger) = \ker(A)^\perp$ and the Moore–Penrose identities*

1. $A^\dagger A = P_{\ker(A)^\perp}$,
2. $AA^\dagger = P_{\overline{\operatorname{im}(A)}}\big|_{\operatorname{dom}(A^\dagger)}$,
3. $AA^\dagger A = A$,
4. $A^\dagger A A^\dagger = A^\dagger$,

where $P_{\ker(A)}$ and $P_{\overline{\operatorname{im}(A)}}$ denote the orthogonal projections onto $\ker(A)$ and $\overline{\operatorname{im}(A)}$, respectively.

Proof. First, by the definition of the Moore–Penrose inverse we have for any $u \in \mathcal{X}$

$$A^\dagger A u = A^\dagger A (P_{\ker(A)} u + P_{\ker(A)^\perp} u) = A^\dagger A P_{\ker(A)^\perp} u = \widetilde{A}^{-1} A P_{\ker(A)^\perp} u = P_{\ker(A)^\perp} u,$$

which proves 1. Now, for any $f \in \operatorname{dom}(A^\dagger)$ we have (see (2.4))

$$AA^\dagger f = A\widetilde{A}^{-1} P_{\overline{\operatorname{im}(A)}} f = P_{\overline{\operatorname{im}(A)}} f,$$

which proves 2. Applying A to 1., we get 3., and applying $A^\dagger$ to 2., we get 4., which completes the proof. □

Corollary 2.15. The Moore–Penrose inverse is uniquely characterised by points 1 and 2 of Theorem 2.14. That is, if a linear operator $B\colon \operatorname{im}(A) \oplus \operatorname{im}(A)^\perp \to \ker(A)^\perp$ satisfies $BA = P_{\ker(A)^\perp}$ and $AB = P_{\overline{\operatorname{im}(A)}}$ then $B = A^\dagger$.

Proof. First we show that $B|_{\mathrm{im}(A)} = \widetilde{A}^{-1}$. Indeed, let $f = Au \in \mathrm{im}(A)$, where $u \in \ker(A)^\perp$. Then

$$Bf = BAu = P_{\ker(A)^\perp} u = u = \widetilde{A}^{-1} f,$$

where the last equality holds since $\widetilde{A}$ is bijective and hence uniquely invertible.

Now we prove that $B|_{\mathrm{im}(A)^\perp} = 0$. Indeed, for any $f \in \mathrm{im}(A)^\perp$ we have

$$ABf = P_{\overline{\mathrm{im}(A)}} f = 0.$$

Since $Bf \in \ker(A)^\perp$ and A is injective on $\ker(A)^\perp$, we conclude that $Bf = 0$. Therefore, B is an extension of $\widetilde{A}^{-1}$ to $\mathrm{im}(A) \oplus \mathrm{im}(A)^\perp$ with $\ker(B) = \mathrm{im}(A)^\perp$. Since such an extension is unique, $B = A^\dagger$. □

Remark 2.16. If an operator B satisfies only $ABA = A$ (resp. $BAB = B$), it is called the *inner inverse* (resp. *outer inverse*) of A.

Let us now characterise when the Moore-Penrose inverse is continuous. We will use the *closed graph theorem* and the following lemma:

Lemma 2.17. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then $A^\dagger$ has a closed graph, i.e. the set* $\mathrm{gr}(A^\dagger) := \{(f, A^\dagger f) | f \in \mathrm{dom}(A^\dagger)\}$ *is closed.*

Proof. Suppose that $(f_n, v_n) = (f_n, A^\dagger f_n)$ is a convergent sequence in $\mathrm{gr}(A^\dagger)$, so that $f_n \to f$ for some $f \in \mathrm{dom}(A^\dagger)$, and $A^\dagger f_n \to v$ for some $v \in \ker(A)^\perp$. By the Moore-Penrose identities, Theorem 2.14, we have

$$Av \leftarrow Av_n = AA^\dagger f_n = P_{\overline{\mathrm{im}(A)}} f_n \to P_{\overline{\mathrm{im}(A)}} f = AA^\dagger f.$$

As a consequence, $Av = AA^\dagger f$, but noting that $v \in \ker(A)^\perp$ and $A^\dagger \in \ker(A)^\perp$, and the fact that A is injective when restricted to $\ker(A)^\perp$, we conclude that $v = A^\dagger f$, implying that $\mathrm{gr}(A)$ is closed. □

Recall one of the important theorems of functional analysis, the closed graph theorem:

Theorem 2.18 (Closed Graph Theorem, [18, Corollary 4.44])**.** *Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and $A : \mathcal{X} \to \mathcal{Y}$ be linear. Then A is continuous if and only if* $\mathrm{gr}(A)$ *is closed.*

With the results above, we are finally equipped to characterise when the Moore–Penrose inverse is continuous:

Theorem 2.19. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then $A^\dagger$ is continuous, i.e. $A^\dagger \in \mathcal{L}(\mathrm{dom}(A^\dagger), \mathcal{X})$, if and only if* $\mathrm{im}(A)$ *is closed.*

Proof. If $\mathrm{im}(A)$ is closed, $\mathrm{dom}(A^\dagger)$ is closed, and hence complete. By Lemma 2.17 and Theorem 2.18, we conclude that $A^\dagger$ is continuous.

If $A^\dagger$ is continuous, it has a unique continuous extension $\overline{A^\dagger}$ to $\mathcal{Y}$, since $\mathrm{dom}(A^\dagger)$ is dense in $\mathcal{Y}$. Furthermore, by the Moore–Penrose identities, Theorem 2.14, we have $A\overline{A^\dagger} = P_{\overline{\mathrm{im}(A)}}$. Taking $f \in \overline{\mathrm{im}(A)}$, we now find that

$$f = P_{\overline{\mathrm{im}(A)}} f = A\overline{A^\dagger} f \in \mathrm{im}(A).$$

As a result, $\overline{\mathrm{im}(A)} \subseteq \mathrm{im}(A)$, implying that $\mathrm{im}(A)$ is closed. □

The next theorem shows that minimum-norm solutions can indeed be computed using the Moore–Penrose generalised inverse.

Theorem 2.20. *For each $f \in \mathrm{dom}(A^\dagger)$, the minimum-norm solution $u^\dagger$ to the inverse problem* (2.1) *is given via*

$$u^\dagger = A^\dagger f.$$

Proof. As $f \in \mathrm{dom}(A^\dagger)$, we know from Theorem 2.9 that the minimum-norm solution $u^\dagger$ exists and is unique. With $u^\dagger \in \ker(A)^\perp$, Lemma 2.14, and Theorem 2.5 we conclude that

$$u^\dagger = P_{\ker(A)^\perp} u^\dagger = A^\dagger A u^\dagger = A^\dagger P_{\overline{\mathrm{im}(A)}} f = A^\dagger A A^\dagger f = A^\dagger f.$$

□

As a consequence of Theorem 2.20 and Theorem 2.5, we find that the minimum-norm solution $u^\dagger$ of $Au = f$ is a minimum-norm solution of the normal equation (2.2), i.e.

$$u^\dagger = (A^*A)^\dagger A^* f.$$

Thus, in order to compute $u^\dagger$ we can equivalently consider finding the minimum-norm solution of the normal equation.

2.2 Compact Operators

Definition 2.21. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then A is said to be* compact *if for any bounded set $B \subset \mathcal{X}$ the closure of its image $\overline{A(B)}$ is compact in $\mathcal{Y}$. We denote the space of compact operators by $\mathcal{K}(\mathcal{X}, \mathcal{Y})$.*

Remark 2.22. We can equivalently define an operator A to be compact if the image of a bounded sequence $\{u_j\}_{j\in\mathbb{N}} \subset \mathcal{X}$ contains a convergent subsequence $\{Au_{j_k}\}_{k\in\mathbb{N}} \subset \mathcal{Y}$.

Compact operators are very common in inverse problems. In fact, almost all (linear) inverse problems involve the inversion of a compact operator. As the following result shows, compactness of the forward operator is a major source of ill-posedness.

Theorem 2.23. *Let $A \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$ have an infinite-dimensional range. Then, the Moore–Penrose inverse of A is discontinuous.*

Proof. As the range $\mathrm{im}(A)$ is of infinite dimension, we can conclude that $\mathcal{X}$ and $\ker(A)^\perp$ are also infinite dimensional. We can therefore find a sequence $\{u_j\}_{j\in\mathbb{N}}$ with $u_j \in \ker(A)^\perp$, $\|u_j\|_\mathcal{X} = 1$ and $\langle u_j, u_k\rangle_\mathcal{X} = 0$ for $j \neq k$. Since A is a compact operator the sequence $f_j = Au_j$ has a convergent subsequence, hence, for all $\delta > 0$ we can find j, k such that $\|f_j - f_k\|_\mathcal{Y} < \delta$. However, we also obtain

$$\begin{aligned}\|A^\dagger f_j - A^\dagger f_k\|_\mathcal{X}^2 &= \|A^\dagger A u_j - A^\dagger A u_k\|_\mathcal{X}^2 \\ &= \|u_j - u_k\|_\mathcal{X}^2 = \|u_j\|_\mathcal{X}^2 - 2\langle u_j, u_k\rangle_\mathcal{X} + \|u_k\|_\mathcal{X}^2 = 2,\end{aligned}$$

which shows that $A^\dagger$ is discontinuous. Here, the second identity follows from Lemma 2.14, point 1, and the fact that $u_j, u_k \in \ker(A)^\perp$. □

To get a better understanding of when we have $f \in \overline{\mathrm{im}(A)} \setminus \mathrm{im}(A)$ for compact operators A, we will consider the singular value decomposition of compact operators.

Singular value decomposition of compact operators

Theorem 2.24 (Spectral theorem for compact operators [18, Theorem 7.34])**.** *Let $\mathcal{X}$ be a Hilbert space and $A \in \mathcal{K}(\mathcal{X}, \mathcal{X})$ be self-adjoint. Then there exists an orthonormal basis $\{x_j\}_{j\in\mathbf{N}} \subset \mathcal{X}$ of $\overline{\operatorname{im}(A)}$ and a sequence of eigenvalues $\{\lambda_j\}_{j\in\mathbf{N}} \subset \mathbf{R}$ with $|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots > 0$ such that for all $u \in \mathcal{X}$ we have*

$$Au = \sum_{j=1}^{\infty} \lambda_j \langle u, x_j\rangle_{\mathcal{X}} x_j.$$

The sequence $\{\lambda_j\}_{j\in\mathbf{N}}$ is either finite or we have $\lambda_j \longrightarrow 0$.

Remark 2.25. The notation in the theorem above only makes sense if the sequence $\{\lambda_j\}_{j\in\mathbf{N}}$ is infinite. For the case that there are only finitely many λ_j the sum has to be interpreted as a finite sum. Moreover, as the eigenvalues are sorted by absolute value $|\lambda_j|$, we have $\|A\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} = |\lambda_1|$.

If A is not self-adjoint, the decomposition in Theorem 2.24 does not hold any more. Instead, we can consider the so-called *singular value decomposition* of a compact linear operator. To prove its existence, we will use the following lemma on the relationship between the ranges of A and AA^*:

Lemma 2.26. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then $\overline{\operatorname{im}(AA^*)} = \overline{\operatorname{im}(A)}$.*

Proof. It is clear that $\overline{\operatorname{im}(AA^*)} = \overline{\operatorname{im}(A|_{\operatorname{im}(A^*)})} \subseteq \overline{\operatorname{im}(A)}$, so we are left to prove that $\overline{\operatorname{im}(A)} \subseteq \overline{\operatorname{im}(AA^*)}$.

Let $f \in \overline{\operatorname{im}(A)}$ and let $\varepsilon > 0$. Then, there exists $u \in \ker(A)^\perp$ with $\|f - Au\|_{\mathcal{X}} < \varepsilon/2$ (recall the orthogonal decomposition in Remark 2.2). As $\ker(A)^\perp = \overline{\operatorname{im}(A^*)}$, there exists $g \in \mathcal{Y}$ such that $\|A^*g - u\|_{\mathcal{X}} < \varepsilon/(2\|A\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})})$. Putting these together we have

$$\begin{aligned} \|AA^*g - f\|_{\mathcal{Y}} &\leqslant \|AA^*g - Au\|_{\mathcal{Y}} + \|Au - f\|_{\mathcal{Y}} \\ &\leqslant \underbrace{\|A\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})}\|A^*g - u\|_{\mathcal{X}}}_{<\varepsilon/2} + \underbrace{\|f - Au\|_{\mathcal{Y}}}_{<\varepsilon/2} < \varepsilon \end{aligned}$$

which shows that $u \in \overline{\operatorname{im}(AA^*)}$ and thus also that $\overline{\operatorname{im}(A)} \subseteq \overline{\operatorname{im}(AA^*)}$. □

Theorem 2.27 (Singular value decomposition)**.** *Let $A \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$. Then there exists*

1. *a not-necessarily infinite null sequence $\{\sigma_j\}_{j\in\mathbf{N}}$ with $\sigma_1 \geqslant \sigma_2 \geqslant \ldots > 0$,*
2. *an orthonormal basis $\{x_j\}_{j\in\mathbf{N}} \subset \mathcal{X}$ of $\ker(A)^\perp$,*
3. *an orthonormal basis $\{y_j\}_{j\in\mathbf{N}} \subset \mathcal{Y}$ of $\overline{\operatorname{im}(A)}$ with*

$$Ax_j = \sigma_j y_j, \quad A^* y_j = \sigma_j x_j, \quad \text{for all } j \in \mathbf{N}. \tag{2.5}$$

Moreover, for all $u \in \mathcal{X}$ we have the representation

$$Au = \sum_{j=1}^{\infty} \sigma_j \langle u, x_j\rangle y_j. \tag{2.6}$$

The sequence $\{(\sigma_j, x_j, y_j)\}$ is called singular system *or* singular value decomposition (SVD) *of A. For the adjoint operator A^* we have the representation*

$$A^* f = \sum_{j=1}^{\infty} \sigma_j \langle f, y_j\rangle x_j \quad \forall f \in \mathcal{Y}. \tag{2.7}$$

Proof. Consider the operator $AA^* \in \mathcal{L}(\mathcal{Y}, \mathcal{Y})$, which is self-adjoint ($(AA^*)* = A^{**}A^* = AA^*$) and compact, being the product of compact operators. Hence, we can apply the spectral theorem, Theorem 2.24, to AA^* to get a sequence of eigenvalues $\{\lambda_j\}_{j\in\mathbf{N}} \subset \mathbf{R}$ (ordered by decreasing absolute value), and corresponding eigenvectors $\{y_j\}_{j\in\mathbf{N}}$, forming an orthonormal basis of $\overline{\operatorname{im}(AA^*)} = \overline{\operatorname{im}(A)}$ (by Lemma 2.26). Then we have

$$\lambda_j = \lambda_j \langle y_j, y_j\rangle = \langle AA^* y_j, y_j\rangle = \|Ay_j\|^2 \geqslant 0,$$

so that we may as well write $\lambda_j = \sigma_j^2$ for $\sigma_j = \sqrt{\lambda_j}$. We also easily see that $x_j := A^* y_j/\sigma_j$ defines an orthonormal system: we have

$$\langle x_j, x_k\rangle = \frac{\langle A^* y_j, A^* y_k\rangle}{\sigma_j\sigma_k} = \frac{\langle AA^* y_j, y_k\rangle}{\sigma_j\sigma_k} = \frac{\sigma_j^2}{\sigma_j\sigma_k}\langle y_j, y_k\rangle = \delta_{jk}$$

with δ_{jk} the Kronecker delta. As a result, $\{x_j\}_{j\in\mathbf{N}}$ is an orthonormal basis for $\overline{\operatorname{span}\{x_j\}_{j\in\mathbf{N}}} = \{x_j\}_{j\in\mathbf{N}}^{\perp\perp}$. To characterise this space, we need only look at its orthogonal complement:

$$\begin{aligned}
v \in \{x_j\}_{j\in\mathbf{N}} &\iff \forall j \in \mathbf{N}, \langle x_j, v\rangle = 0 \\
&\iff \forall j \in \mathbf{N}, \langle y_j, Av\rangle = 0 \\
&\iff Av \in \operatorname{im}(A)^\perp && (\{y_j\}_{j\in\mathbf{N}} \text{ is a basis for } \overline{\operatorname{im}(A)}) \\
&\iff Av = 0.
\end{aligned}$$

We conclude that $\{x_j\}_{j\in\mathbf{N}}$ is an orthonormal basis of $\ker(A)^\perp = \overline{\operatorname{im}(A^*)}$

Furthermore, we have (by definition of x_j and the fact that y_j is an eigenvector of AA^*)

$$A^* y_j = \sigma_j x_j, \quad Ax_j = \frac{1}{\sigma_j} AA^* y_j = \sigma_j y_j.$$

We are now ready to put everything together and conclude. Since $\{x_j\}_{j\in\mathbf{N}}$ is an orthonormal basis for $\overline{\operatorname{im}(A^*)}$ and $\{y_j\}_{j\in\mathbf{N}}$ is an orthonormal basis for $\overline{\operatorname{im}(A)}$, we have the following identities, for $u \in \mathcal{X}$ and $f \in \mathcal{Y}$:

$$P_{\overline{\operatorname{im}(A^*)}} u = \sum_{j\in\mathbf{N}} \langle u, x_j\rangle x_j, \quad P_{\overline{\operatorname{im}(A)}} f = \sum_{j\in\mathbf{N}} \langle f, y_j\rangle y_j.$$

In particular, we have

$$\begin{aligned}
Au &= A(P_{\ker(A)} u + P_{\overline{\operatorname{im}(A^*)}} u) \\
&= AP_{\overline{\operatorname{im}(A^*)}} u \\
&= A\Big(\sum_{j\in\mathbf{N}} \langle u, x_j\rangle x_j\Big) \\
&= \sum_{j\in\mathbf{N}} \langle u, x_j\rangle Ax_j = \sum_{j\in\mathbf{N}} \sigma_j \langle u, x_j\rangle y_j,
\end{aligned}$$

and similarly,

$$\begin{aligned}
A^* f &= A^*(P_{\ker(A^*)} f + P_{\overline{\operatorname{im}(A)}} f) \\
&= A^* P_{\overline{\operatorname{im}(A)}} f \\
&= A^*\Big(\sum_{j\in\mathbf{N}} \langle f, y_j\rangle y_j\Big) \\
&= \sum_{j\in\mathbf{N}} \langle f, y_j\rangle A^* y_j = \sum_{j\in\mathbf{N}} \sigma_j \langle f, y_j\rangle x_j.
\end{aligned}$$

□

We can now derive a representation of the Moore–Penrose inverse in terms of the singular value decomposition.

Theorem 2.28. *Let $A \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$ with singular system $\{(\sigma_j, x_j, y_j)\}_{j\in\mathbb{N}}$ and $f \in \operatorname{dom}(A^\dagger)$. Then the Moore–Penrose inverse of A can be written as*

$$A^\dagger f = \sum_{j=1}^{\infty} \sigma_j^{-1} \langle f, y_j \rangle x_j. \tag{2.8}$$

Proof. We know that, since $f \in \operatorname{dom}(A^\dagger)$, $u^\dagger = A^\dagger f$ solves the normal equation

$$A^* A u^\dagger = A^* f.$$

From Theorem 2.27 we know that

$$A^* A u^\dagger = \sum_{j=1}^{\infty} \sigma_j^2 \langle u^\dagger, x_j \rangle x_j, \quad A^* f = \sum_{j=1}^{\infty} \sigma_j \langle f, y_j \rangle x_j, \tag{2.9}$$

which implies that

$$\langle u^\dagger, x_j \rangle = \sigma_j^{-1} \langle f, y_j \rangle$$

Expanding $u^\dagger \in \ker(A)^\perp$ in the basis $\{x_j\}$, we get

$$u^\dagger = \sum_{j=1}^{\infty} \langle u^\dagger, x_j \rangle x_j = \sum_{j=1}^{\infty} \sigma_j^{-1} \langle f, y_j \rangle x_j = A^\dagger f.$$

□

The representation (2.8) makes it clear again that the Moore–Penrose inverse is unbounded if $\operatorname{im}(A)$ is infinite dimensional. Indeed, taking the sequence y_j we note that $\|A^\dagger y_j\| = \sigma_j^{-1} \to \infty$, although $\|y_j\| = 1$. The unboundedness of the Moore–Penrose inverse is also reflected in the fact that the series in (2.8) may not converge for a given f. The convergence criterion for the series is called the *Picard criterion*:

Definition 2.29. *We say that the data f satisfy the Picard criterion, if*

$$\|A^\dagger f\|^2 = \sum_{j=1}^{\infty} \frac{\left|\langle f, y_j \rangle\right|^2}{\sigma_j^2} < \infty. \tag{2.10}$$

Remark 2.30. The Picard criterion is a condition on the decay of the coefficients $\langle f, y_j \rangle$. As the singular values σ_j decay to zero as $j \to \infty$, the Picard criterion is only met if the coefficients $\langle f, y_j \rangle$ decay sufficiently fast. In case the singular system is given by the Fourier basis, then the coefficients $\langle f, y_j \rangle$ are just the Fourier coefficients of f. Therefore, the Picard criterion is a condition on the decay of the Fourier coefficients which is equivalent to the smoothness of f.

The Picard criterion gives us another way to characterise elements in the range of the forward operator:

Theorem 2.31. *Let $A \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$ with singular system $\{(\sigma_j, x_j, y_j)\}_{j\in\mathbb{N}}$, and $f \in \overline{\operatorname{im}(A)}$. Then $f \in \operatorname{im}(A)$ if and only if the* Picard criterion

$$\sum_{j=1}^{\infty} \frac{|\langle f, y_j \rangle_{\mathcal{Y}}|^2}{\sigma_j^2} < \infty$$

is met.

Proof. Let $f \in \operatorname{im}(A)$, so that there is a $u \in \mathcal{X}$ such that $Au = f$. It is easy to see that we have

$$\langle f, y_j \rangle_{\mathcal{Y}} = \langle Au, y_j \rangle_{\mathcal{Y}} = \langle u, A^* y_j \rangle_{\mathcal{X}} = \sigma_j \langle u, x_j \rangle_{\mathcal{X}}$$

and therefore

$$\sum_{j=1}^{\infty} \frac{|\langle f, y_j \rangle_{\mathcal{Y}}|^2}{\sigma_j^2} = \sum_{j=1}^{\infty} |\langle u, x_j \rangle_{\mathcal{X}}|^2 \leqslant \|u\|_{\mathcal{X}}^2 < \infty.$$

Now let the Picard criterion (2.10) hold and define $u := \sum_{j=1}^{\infty} \sigma_j^{-1} \langle f, y_j \rangle_{\mathcal{Y}} x_j \in \mathcal{X}$. It is well-defined by the Picard criterion (2.10) and we conclude

$$Au = \sum_{j=1}^{\infty} \sigma_j^{-1} \langle f, y_j \rangle_{\mathcal{Y}} A x_j = \sum_{j=1}^{\infty} \langle f, y_j \rangle_{\mathcal{Y}} y_j = P_{\overline{\operatorname{im}(A)}} f = f,$$

which shows that $f \in \operatorname{im}(A)$. □

Although all ill-posed problems are not easy to solve, some are worse than others, depending on how fast the singular values decay to zero.

Definition 2.32. *We say that an ill-posed inverse problem* (2.1) *is* mildly ill-posed *if the singular values decay at most with polynomial speed, i.e. there exist $\gamma, C > 0$ such that $\sigma_j \geqslant C j^{-\gamma}$ for all j. We call the ill-posed inverse problem* severely ill-posed *if its singular values decay faster than with polynomial speed, i.e. for all $\gamma, C > 0$ one has that $\sigma_j \leqslant C j^{-\gamma}$ for j sufficiently large.*

Example 2.33. Let us consider the example of differentiation again, as introduced in Section 1.2.3. The forward operator $A \colon L^2([0,1]) \to L^2([0,1])$ in this problem is given by

$$(Au)(t) = \int_0^t u(s)\,\mathrm{d}s = \int_0^1 K(s,t) u(s)\,\mathrm{d}s,$$

with $K \colon [0,1] \times [0,1] \to \mathbf{R}$ defined as

$$K(s,t) := \begin{cases} 1 & s \leqslant t, \\ 0 & \text{else.} \end{cases}$$

This is a special case of the integral operators as introduced in Section 1.2.1. Since the kernel K is square integrable, A is compact. The adjoint operator A^* is given by

$$(A^* f)(s) = \int_0^1 K(t,s) f(t)\,\mathrm{d}t = \int_s^1 v(t)\,\mathrm{d}t. \tag{2.11}$$

Now we want to compute the eigenvalues and eigenvectors of A^*A, i.e. we look for σ^2 and $x \in L^2([0,1])$ with

$$\sigma^2 x(s) = (A^* A x)(s) = \int_s^1 \int_0^t x(r)\,\mathrm{d}r\,\mathrm{d}t.$$

We immediately observe that $x(1) = 0$ and further

$$\sigma^2 x'(s) = \frac{\mathrm{d}}{\mathrm{d}s} \int_s^1 \int_0^t x(r)\,\mathrm{d}r\,\mathrm{d}t = -\int_0^s x(r)\,\mathrm{d}r,$$

from which we conclude $x'(0) = 0$. Taking the derivative another time yields the ordinary differential equation

$$\sigma^2 x''(s) + x(s) = 0,$$

for which solutions are of the form

$$x(s) = c_1 \sin(\sigma^{-1}s) + c_2 \cos(\sigma^{-1}s),$$

with some constants c_1, c_2. In order to satisfy the boundary conditions $x(1) = c_1 \sin(\sigma^{-1}) + c_2 \cos(\sigma^{-1}) = 0$ and $x'(0) = c_1/\sigma = 0$, we choose $c_1 = 0$ and σ such that $\cos(\sigma^{-1}) = 0$. Hence, we have

$$\sigma_j = \frac{2}{(2j-1)\pi} \text{ for } j \in \mathbf{N},$$

and by choosing $c_2 = \sqrt{2}$ we obtain the following normalised representation of x_j:

$$x_j(s) = \sqrt{2} \cos\left(\left(j - \frac{1}{2}\right)\pi s\right).$$

According to (2.5) we further obtain

$$y_j(s) = \sigma_j^{-1}(Ax_j)(s) = \left(j - \frac{1}{2}\right)\pi \int_0^s \sqrt{2} \cos\left(\left(j - \frac{1}{2}\right)\pi t\right) \mathrm{d}t = \sqrt{2} \sin\left(\left(j - \frac{1}{2}\right)\pi s\right),$$

and hence, for $f \in L^2([0,1])$ the Picard criterion becomes

$$2 \sum_{j=1}^{\infty} \sigma_j^{-2} \left(\int_0^1 f(s) \sin\left(\sigma_j^{-1}s\right)\,\mathrm{d}s\right)^2 < \infty.$$

Expanding f in the basis $\{y_j\}$

$$f(t) = 2 \sum_{j=1}^{\infty} \left(\int_0^1 f(s) \sin\left(\sigma_j^{-1}s\right)\,\mathrm{d}s\right) \sin\left(\sigma_j^{-1}t\right),$$

and formally differentiating the series, we obtain

$$f'(t) = 2 \sum_{j=1}^{\infty} \sigma_j^{-1} \left(\int_0^1 f(s) \sin\left(\sigma_j^{-1}s\right)\,\mathrm{d}s\right) \cos\left(\sigma_j^{-1}t\right).$$

Therefore, the Picard criterion is nothing but the condition for the legitimacy of such differentiation, i.e. for the differentiability of the Fourier series by differentiating its components, and it holds if f is differentiable and $f' \in L^2([0,1])$.From the decay of the singular values we see that this inverse problem is mildly ill posed.

Example 2.34 (Heat equation)**.** Consider the problem of recovering the initial condition u of the heat equation from an observation f of the solution at some time $T > 0$ (see Section 1.2.2). We consider the heat equation on $(0, \pi) \times \mathbf{R}_+$, with Dirichlet boundary conditions

$$\begin{cases} v_t - v_{xx} = 0 & \text{on } (0,\pi) \times \mathbf{R}_+, \\ v(0,t) = v(\pi,t) = 0 & \text{on } \mathbf{R}_+, \\ v(x,T) = f(x) & \text{on } (0,\pi), \\ v(x,0) = u(x) & \text{on } (0,\pi). \end{cases}$$

The solution to the forward problem (determine f given u) is given by

$$f = Au := \sum_{j=1}^{\infty} \exp(-j^2 T) \widehat{u}_j \sin(jx),$$

where $\widehat{u}_j = \langle u, \sin(j\cdot)\rangle$ are Fourier coefficients of u. Hence, singular values of A are given by

$$\sigma_j = \exp(-j^2 T), \quad j \in \mathbf{N},$$

and

$$\frac{1}{\sigma_j} = \exp(j^2 T).$$

The singular values of A decay exponentially and the inverse problem is severely (exponentially) ill-posed.

Chapter 3

Classical Regularisation Theory

3.1 What is Regularisation?

We have seen that the Moore–Penrose inverse $A^\dagger$ is unbounded if $\mathrm{im}(A)$ is not closed. Therefore, given noisy data f_δ such that $\|f_\delta - f\| \leqslant \delta$, we cannot expect convergence $A^\dagger f_\delta \to A^\dagger f$ as $\delta \to 0$. To achieve convergence, we replace $A^\dagger$ with a family of well-posed (bounded) operators R_α with $\alpha = \alpha(\delta, f_\delta)$ and require that $R_{\alpha(\delta,f_\delta)}(f_\delta) \to A^\dagger f$ for all $f \in \mathrm{dom}(A^\dagger)$ and all $f_\delta \in \mathcal{Y}$ s.t. $\|f - f_\delta\|_\mathcal{Y} \leqslant \delta$ as $\delta \to 0$. We remind ourselves that $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ denotes the space of all bounded (equivalently, continuous) linear operators $\mathcal{X} \to \mathcal{Y}$.

Definition 3.1. *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be a bounded operator. A family $\{R_\alpha\}_{\alpha>0}$ of continuous operators is called* regularisation *(or* regularisation operator*) of $A^\dagger$ if*

$$R_\alpha f \to A^\dagger f = u^\dagger$$

for all $f \in \mathrm{dom}(A^\dagger)$ as $\alpha \to 0$.

Definition 3.2. *If the family $\{R_\alpha\}_{\alpha>0}$ consists of linear operators, then one speaks of* linear regularisation *of $A^\dagger$.*

Hence, a regularisation is a pointwise approximation of the Moore–Penrose inverse with continuous operators. In this chapter, we will only concern ourselves with linear regularisations. In most cases of interest, the Moore–Penrose inverse is not continuous, so that we cannot expect that the norm of R_α stays bounded as $\alpha \to 0$. This is confirmed by the following results (in the linear case). For this let us first recall the uniform boundedness principle, also known as the Banach–Steinhaus theorem:

Theorem 3.3 (Uniform Boundedness Principle [18, Theorem 4.52])**.** *Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and $\mathcal{F} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y})$ a family of point-wise bounded operators, i.e. for all $u \in \mathcal{X}$ there exists a constant $C(u) > 0$ s.t. $\sup_{A\in\mathcal{F}} \|Au\|_\mathcal{Y} \leqslant C(u)$. Then*

$$\sup_{A\in\mathcal{F}} \|A\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} < \infty.$$

Corollary 3.4 ([23, p. 174])**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then the following two conditions are equivalent:

1. There exists $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ such that

$$Au = \lim_{j\to\infty} A_j u \quad \text{for all } u \in \mathcal{X}.$$

2. There is a dense subset $\mathcal{X}' \subset \mathcal{X}$ such that $\lim_{j\to\infty} A_j u$ exists for all $u \in \mathcal{X}'$ and

$$\sup_{j\in\mathbb{N}} \|A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} < \infty.$$

Theorem 3.5. *Let $\mathcal{X}$, $\mathcal{Y}$ be Hilbert spaces, $A \in \mathcal{L}(\mathcal{X},\mathcal{Y})$ and $\{R_\alpha\}_{\alpha>0}$ a linear regularisation as defined in Definition 3.2. If $A^\dagger$ is not continuous, $\{R_\alpha\}_{\alpha>0}$ cannot be uniformly bounded. In particular, there exist $f \in \mathcal{Y}$ and a sequence $\alpha_j \to 0$ such that $\|R_{\alpha_j} f\| \to \infty$ as $j \to \infty$.*

Proof. For the first statement, we will prove the contrapositive: assume that $\{R_\alpha\} > 0$ is uniformly bounded. As a consequence, using that $R_\alpha f \to A^\dagger$ for all $f \in \operatorname{dom}(A^\dagger)$ (which is dense in $\mathcal{Y}$), we find by Corollary 3.4 that there is a $B \in \mathcal{L}(\mathcal{X},\mathcal{Y})$ such that $R_\alpha f \to Bf$ for all $f \in \mathcal{Y}$. Of course, then $B|_{\operatorname{dom}(A^\dagger)} = A^\dagger$. Furthermore, by the Moore–Penrose identities (Theorem 2.14), we have $AB|_{\operatorname{dom}(A^\dagger)} = P_{\overline{\operatorname{im}(A)}}|_{\operatorname{dom}(A^\dagger)}$, and continuous functions are uniquely determined by their behaviour on dense sets, so $AB = P_{\overline{\operatorname{im}(A)}}$. In particular, for any $f \in \overline{\operatorname{im}(A)}$, we have $f = P_{\overline{\operatorname{im}(A)}} f = ABf \in \operatorname{im}(A)$, so that $\operatorname{im}(A)$ is closed, and $A^\dagger$ is continuous, by Theorem 2.19.

For the second statement, assume that $A^\dagger$ is discontinuous, and assume that for all $f \in \mathcal{Y}$ and any sequence $\alpha_j \to 0$ we have

$$\sup_{j\in\mathbb{N}} \|R_{\alpha_j} f\|_{\mathcal{Y}} \leqslant C(f) < \infty.$$

Then by Theorem 3.3 we have that

$$\sup_{j\in\mathbb{N}} \|R_{\alpha_j}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \leqslant C < \infty,$$

which contradicts the first part of the proof. □

With the additional assumption that $\|AR_\alpha\|_{\mathcal{L}(\mathcal{X},\mathcal{X})}$ is bounded, we can even show that $R_\alpha f$ diverges for all $f \notin \operatorname{dom}(A^\dagger)$. For this purpose, let us first recall the definition of weak convergence in Hilbert spaces, and some of its consequences. A sequence $\{u_j\}_{j\in\mathbb{N}} \subseteq \mathcal{X}$ is said to converge weakly to $u \in \mathcal{X}$ if it is bounded and $\langle u_j - u, v\rangle \to 0$ for all $v \in \mathcal{X}$. We denote this convergence by $u_n \rightharpoonup u$. It is an elementary calculation to show that if $A \in \mathcal{L}(\mathcal{X},\mathcal{Y})$, then A is weak-to-weak continuous, meaning that $u_n \rightharpoonup u$ implies that $Au_n \rightharpoonup Au$. Finally, any bounded sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{X}$ has a weakly convergent subsequence [18, Theorem 5.73].

Theorem 3.6. *Let $A \in \mathcal{L}(\mathcal{X},\mathcal{Y})$ and $\{R_\alpha\}_{\alpha>0}$ be a linear regularisation of $A^\dagger$. If*

$$\sup_{\alpha>0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} < \infty,$$

then $\|R_\alpha f\|_{\mathcal{X}} \to \infty$ for all $f \notin \operatorname{dom}(A^\dagger)$.

Proof. Define $u_\alpha := R_\alpha f$ for $f \notin \operatorname{dom}(A^\dagger)$. Assume that there exists a sequence $\alpha_k \to 0$ such that $\|u_{\alpha_k}\|_{\mathcal{X}}$ is uniformly bounded. As a consequence, there is a weakly convergent subsequence $u_{\alpha_{k_l}}$ with some limit $u \in \mathcal{X}$. As continuous linear operators are also weak-to-weak continuous, we furthermore have that $Au_{\alpha_{k_l}} \rightharpoonup Au$.

On the other hand, for any $g \in \operatorname{dom}(A^\dagger)$ we have that $AR_{\alpha_{k_l}} g \to AA^\dagger g = P_{\overline{\operatorname{im}(A)}} g$ as $l \to \infty$. By Corollary 3.4 we then conclude that this also holds for any $f \in \mathcal{Y}$, i.e. also for $f \notin \operatorname{dom}(A^\dagger)$. Hence, we get that

$$AR_{\alpha_{k_l}} f \to P_{\overline{\operatorname{im}(A)}} f$$

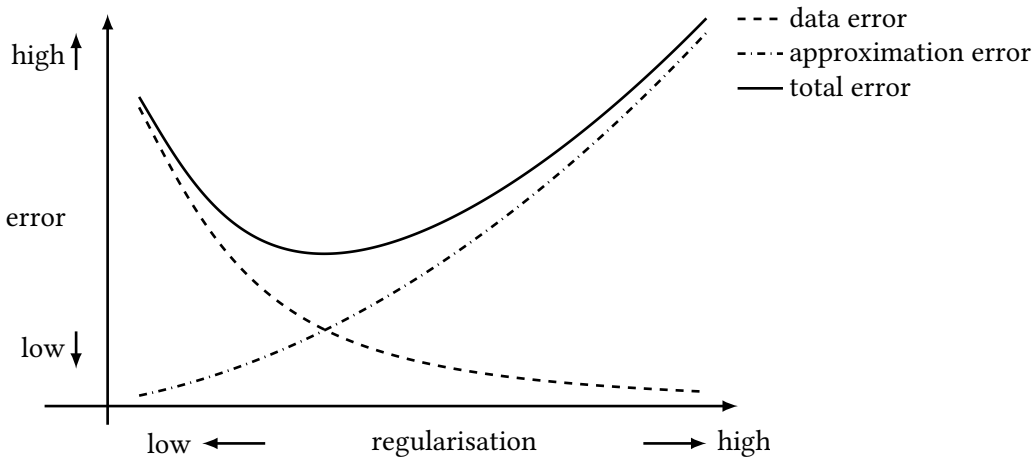

Figure 3.1: The *total error* between a regularised solution and the minimal norm solution decomposes into the *data error* and the *approximation error*. These two errors have opposing trends: For a small regularisation parameter α the error in the data gets amplified through the ill-posedness of the problem and for large α the operator R_α is a poor approximation of the Moore–Penrose inverse.

and (see first part of proof)

$$AR_{\alpha_{k_l}} f = Au_{\alpha_{k_l}} \rightharpoonup Au.$$

Therefore, we get that $Au = P_{\overline{\mathrm{im}(A)}} f$. Since $\mathcal{Y} = \overline{\mathrm{im}(A)} \oplus \mathrm{im}(A)^\perp$, we get that $f \in \mathrm{im}(A) \oplus \mathrm{im}(A)^\perp = \mathrm{dom}(A^\dagger)$ in contradiction to the assumption $f \notin \mathrm{dom}(A^\dagger)$. □

3.2 Parameter Choice Rules

We have stated in the beginning of this chapter that we would like to obtain a regularisation that would guarantee that $R_\alpha(f_\delta) \to A^\dagger f$ for all $f \in \mathrm{dom}(A^\dagger)$ and all $f_\delta \in \mathcal{Y}$ s.t. $\|f - f_\delta\|_\mathcal{Y} \leqslant \delta$ as $\delta \to 0$. This means that the parameter α, referred to as the *regularisation parameter*, needs to be chosen as a function of δ (and perhaps also f_δ) so that $\alpha \to 0$ as $\delta \to 0$ (i.e. we need to regularise less as the data get more precise).

This can be illustrated with the following observation. For linear regularisations we can split the *total error* between the regularised solution of the noisy problem $R_\alpha f_\delta$ and the minimal norm solution of the noise-free problem $u^\dagger = A^\dagger f$ as

$$\begin{aligned} \|R_\alpha f_\delta - u^\dagger\|_\mathcal{X} &\leqslant \|R_\alpha f_\delta - R_\alpha f\|_\mathcal{X} + \|R_\alpha f - u^\dagger\|_\mathcal{X} \\ &\leqslant \underbrace{\delta \|R_\alpha\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})}}_{\text{data error}} + \underbrace{\|R_\alpha f - A^\dagger f\|_\mathcal{X}}_{\text{approximation error}}. \end{aligned} \tag{3.1}$$

The first term of (3.1) is the *data error*; this term unfortunately does not stay bounded for $\alpha \to 0$, which we can conclude from Theorem 3.5. The second term, known as the *approximation error*, however vanishes for $\alpha \to 0$, due to the pointwise convergence of R_α to $A^\dagger$. Hence it becomes evident from (3.1) that a good choice of α depends on δ, and needs to be chosen such that the approximation error becomes as small as possible, whilst the data error is being kept at bay. See Figure 3.1 for an illustration. Parameter choice rules are defined as follows.

Definition 3.7. *A function* $\alpha\colon \mathbf{R}_{>0}\times\mathcal{Y}\to\mathbf{R}_{>0}$, $(\delta, f_\delta)\mapsto\alpha(\delta, f_\delta)$ *is called a* parameter choice rule. *We distinguish between*

1. *a-priori parameter choice rules, which depend on* δ *only;*
2. *a-posteriori parameter choice rules, which depend on both* δ *and* f_δ*;*
3. *heuristic parameter choice rules, which depend on* f_δ *only.*

Now we are ready to define a regularisation that ensures the convergence $R_{\alpha(\delta,f_\delta)}(f_\delta)\to A^\dagger f$ as $\delta\to 0$.

Definition 3.8. *Let* $\{R_\alpha\}_{\alpha>0}$ *be a regularisation of* $A^\dagger$. *If there exists a parameter choice rule* $\alpha : \mathbf{R}_{>0}\times\mathcal{Y}\to\mathbf{R}_{>0}$ *such that for all* $f\in\mathrm{dom}(A^\dagger)$

$$\lim_{\delta\to 0}\sup_{f_\delta\colon \|f-f_\delta\|_{\mathcal{Y}}\leqslant\delta}\|R_{\alpha(\delta,f_\delta)}f_\delta - A^\dagger f\|_{\mathcal{X}} = 0 \tag{3.2}$$

and

$$\lim_{\delta\to 0}\sup_{f_\delta\colon \|f-f_\delta\|_{\mathcal{Y}}\leqslant\delta}\alpha(\delta, f_\delta) = 0 \tag{3.3}$$

then the pair $(\{R_\alpha\}_{\alpha>0},\alpha)$ *is called a* convergent regularisation.

3.2.1 A priori parameter choice rules

To start off, we will discuss a-priori parameter choice rules in more detail. Historically, they were the first to be studied. For every regularisation there exists an a-priori parameter choice rule and thus a convergent regularisation.

Theorem 3.9 ([12, Proposition 3.4])**.** *Let* $\{R_\alpha\}_{\alpha>0}$ *be a regularisation of* $A^\dagger$, *for* $A\in\mathcal{L}(\mathcal{X},\mathcal{Y})$. *Then there exists an a-priori parameter choice rule* $\alpha=\alpha(\delta)$ *such that* $(\{R_\alpha\}_{\alpha>0},\alpha)$ *is a convergent regularisation.*

For linear regularisations, an important characterisation of a-priori parameter choice strategies that lead to convergent regularisation methods is as follows.

Theorem 3.10. *Let* $\{R_\alpha\}_{\alpha>0}$ *be a linear regularisation, and* $\alpha : \mathbf{R}_{>0}\to\mathbf{R}_{>0}$ *an a-priori parameter choice rule. Then* $(\{R_\alpha\}_{\alpha>0},\alpha)$ *is a convergent regularisation method if and only if*

a) $\lim_{\delta\to 0}\alpha(\delta) = 0$

b) $\lim_{\delta\to 0}\delta\|R_{\alpha(\delta)}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} = 0$

Proof. $\Leftarrow$: Let condition a) and b) be fulfilled. From (3.1) we then observe that for any $f\in\mathrm{dom}(A^\dagger)$ and $f_\delta\in\mathcal{Y}$ s.t. $\|f-f_\delta\|_{\mathcal{Y}}\leqslant\delta$

$$\|R_{\alpha(\delta)}f_\delta - A^\dagger f\|_{\mathcal{X}}\to 0 \text{ for } \delta\to 0.$$

Hence, $(\{R_\alpha\}_{\alpha>0},\alpha)$ is a convergent regularisation method.
$\Rightarrow$: Now let $(\{R_\alpha\}_{\alpha>0},\alpha)$ be a convergent regularisation method. We prove that conditions 1 and 2 have to follow from this by showing that violation of either one of them leads to a contradiction to $(\{R_\alpha\}_{\alpha>0},\alpha)$ being a convergent regularisation method. If condition a) is violated, (3.3) is violated and hence, $(\{R_\alpha\}_{\alpha>0},\alpha)$ is not a convergent regularisation method. If condition a) is fulfilled but

condition b) is violated, there exists a null sequence $\{\delta_k\}_{k\in\mathbf{N}}$ with $\delta_k\|R_{\alpha(\delta_k)}\|_{\mathcal{L}(\mathcal{Y},\mathcal{X})} \geqslant C > 0$, and hence, we can find a sequence $\{g_k\}_{k\in\mathbf{N}} \subset \mathcal{Y}$ with $\|g_k\|_{\mathcal{Y}} = 1$ and $\delta_k\|R_{\alpha(\delta_k)}g_k\|_{\mathcal{X}} \geqslant \widetilde{C}$ for some $\widetilde{C}$. Let $f \in \mathrm{dom}(A^\dagger)$ be arbitrary and define $f_k := f + \delta_k g_k$. Then we have on the one hand $\|f - f_k\|_{\mathcal{Y}} \leqslant \delta_k$, but on the other hand the norm of

$$R_{\alpha(\delta_k)}f_k - A^\dagger f = R_{\alpha(\delta_k)}f - A^\dagger f + \delta_k R_{\alpha(\delta_k)}g_k$$

cannot converge to zero, as the second term $\delta_k R_{\alpha(\delta_k)}g_k$ is bounded from below by a positive constant C by construction. Hence, (3.2) is violated for $f_\delta = f + \delta_k g_k$ and thus, $(\{R_\alpha\}_{\alpha>0}, \alpha)$ is not a convergent regularisation method. □

3.2.2 A posteriori parameter choice rules

It is easy to convince oneself that if an a-priori parameter choice rule $\alpha = \alpha(\delta)$ defines a convergence regularisation then $\widetilde{\alpha} = \alpha(C\delta)$ with any $C > 0$ also defines a convergent regularisation (for linear regularisations, it is a trivial corollary of Theorem 3.10). Therefore, from the asymptotic point of view, all these regularisations are equivalent. For a fixed error level δ, however, they can produce very different solutions. Since in practice we have to deal with a typically small, but fixed δ, we would like to have a parameter choice rule that is sensitive to this value. To achieve this, we need to use more information than merely the error level δ to choose the parameter α and we will obtain this information from the approximate data f_δ.

The basic idea is as follows. Let $f \in \mathrm{dom}(A^\dagger)$ and $f_\delta \in \mathcal{Y}$ such that $\|f - f_\delta\| \leqslant \delta$ and consider the *residual* between f_δ and $u_\alpha := R_\alpha f_\delta$, i.e.

$$\|Au_\alpha - f_\delta\|.$$

Let $u^\dagger$ be the minimal norm solution and define

$$\mu := \inf\{\|Au - f\| | u \in \mathcal{X}\} = \|Au^\dagger - f\|.$$

We observe that $u^\dagger$ satisfies the following inequality

$$\|Au^\dagger - f_\delta\| \leqslant \|Au^\dagger - f\| + \|f_\delta - f\| \leqslant \mu + \delta$$

and in some cases this estimate may be sharp. Hence, it appears not to be useful to choose $\alpha(\delta, f_\delta)$ with $\|Au_\alpha - f_\delta\| < \mu + \delta$. In general, it may be not straightforward to estimate μ, but if $\mathrm{im}(A)$ is dense in $\mathcal{Y}$, we get that $\mathrm{im}(A)^\perp = \{0\}$ due to Remark 2.2 and $\mu = 0$. Therefore, we ideally ensure that $\mathrm{im}(A)$ is dense.

These observations motivate the Morozov's discrepancy principle, which in the case $\mu = 0$ reads as follows.

Definition 3.11 (Morozov's discrepancy principle)**.** *Let $u_\alpha = R_\alpha f_\delta$ with $\alpha(\delta, f_\delta)$ chosen as follows*

$$\alpha(\delta, f_\delta) = \sup\{\alpha > 0 \mid \|Au_\alpha - f_\delta\| \leqslant \eta\delta\} \tag{3.4}$$

for given δ, f_δ and a fixed constant $\eta > 1$. Then $u_{\alpha(\delta,f_\delta)} = R_{\alpha(\delta,f_\delta)}f_\delta$ is said to satisfy Morozov's discrepancy principle.

It can be shown that the a-posteriori parameter choice rule (3.4) indeed yields a convergent regularisation method [12, Chapter 4.3].

3.2.3 Heuristic parameter choice rules

As the measurement error δ is not always easy to obtain in practice, it is tempting to use a parameter choice rule that only depends on the measured data f_δ and not on their error δ, i.e. to use a heuristic parameter choice rule. Unfortunately, heuristic rules can yield convergent regularisations only for well-posed problems, as the following result, known as the Bakushinskii veto [5], demonstrates.

Theorem 3.12 ([12, Theorem 3.3])**.** *Let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ and $\{R_\alpha\}$ be a regularisation for $A^\dagger$. Suppose that $\alpha = \alpha(f_\delta)$ be a parameter choice rule such that $(\{R_\alpha\}_{\alpha>0}, \alpha)$ is a convergent regularisation. Then the inverse problem under consideration is in fact* well-posed*, i.e. $A^\dagger$ is continuous.*

Proof. Let $f \in \operatorname{dom}(A^\dagger)$. For any δ, we have $\|f - f\| = 0 \leqslant \delta$, so

$$0 \leqslant \|R_{\alpha(f)} f - A^\dagger f\| \leqslant \sup_{f_\delta : \|f - f_\delta\|_{\mathcal{Y}} \leqslant \delta} \|R_{\alpha(f_\delta)} f_\delta - A^\dagger f\|.$$

By assumption, the right hand side converges to 0 as $\delta \to 0$, so we conclude that $\|R_{\alpha(f)} f - A^\dagger f\| = 0$, or $R_{\alpha(f)} f = A^\dagger f$. Now let $\{f_n\}_{n \in \mathbf{N}} \subset \operatorname{dom}(A^\dagger)$ be any sequence converging to f. By the same argument, we have $R_{\alpha(f_n)} = A^\dagger f_n$, but since $f_n \to f$ and we have assumed that $(\{R_\alpha\}_{\alpha>0}, \alpha)$ is a convergent regularisation, we see that

$$A^\dagger f_n = R_{\alpha(f_n)} f_n \to R_{\alpha(f)} f \to A^\dagger f.$$

We conclude that $A^\dagger$ is continuous, meaning in particular (recall Theorem 2.17) that $\operatorname{im}(A)$ is closed. □

3.3 Spectral Regularisation

In this section, we will return to the setting of compact forward operators $A \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$. Recall that the SVD of A then gives us the spectral representation (2.8) of the Moore–Penrose inverse $A^\dagger$ (Theorem 2.28)):

$$A^\dagger f = \sum_{j=1}^{\infty} \frac{1}{\sigma_j} \langle f, y_j \rangle x_j,$$

where $\{(\sigma_j, x_j, y_j)\}$ is the singular system of A The source of ill-posedness of $A^\dagger$ are the singular values $1/\sigma_j$, which explode as $j \to \infty$, since $\sigma_j \to 0$ as $j \to \infty$. Let us construct a regularisation by modifying these singular values as follows

$$R_\alpha f := \sum_{j=1}^{\infty} g_\alpha(\sigma_j) \langle f, y_j \rangle x_j, \quad f \in \mathcal{Y}, \tag{3.5}$$

with an appropriate function $g_\alpha \colon \mathbf{R}_+ \to \mathbf{R}_+$ such that $g_\alpha(\sigma) \to \frac{1}{\sigma}$ as $\alpha \to 0$ for all $\sigma > 0$ and

$$g_\alpha(\sigma) \leqslant C_\alpha \text{ for all } \sigma \in \mathbf{R}_+. \tag{3.6}$$

Theorem 3.13. *Let $g_\alpha : \mathbf{R}_+ \to \mathbf{R}_+$ be a piecewise continuous function satisfying* (3.6)*, $\lim_{\alpha \to 0} g_\alpha(\sigma) = \frac{1}{\sigma}$ and*

$$\sup_{\alpha, \sigma} \sigma g_\alpha(\sigma) \leqslant \gamma \tag{3.7}$$

for some constant $\gamma > 0$. If R_α is defined as in (3.5)*, we have*

$$R_\alpha f \to A^\dagger f \text{ as } \alpha \to 0$$

for all $f \in \mathrm{dom}(A^\dagger)$.

Proof. From the singular value decomposition of $A^\dagger$ and the definition of R_α we obtain

$$R_\alpha f - A^\dagger f = \sum_{j=1}^{\infty} \Big(g_\alpha(\sigma_j) - \frac{1}{\sigma_j}\Big)\langle f, y_j\rangle_Y x_j = \sum_{j=1}^{\infty}(\sigma_j g_\alpha(\sigma_j) - 1)\langle u^\dagger, x_j\rangle_X x_j.$$

Consider

$$\|R_\alpha f - A^\dagger f\|_X^2 = \sum_{j=1}^{\infty}(\sigma_j g_\alpha(\sigma_j) - 1)^2 |\langle u^\dagger, x_j\rangle_X|^2.$$

From (3.7) we can conclude

$$(\sigma_j g_\alpha(\sigma_j) - 1)^2 \leqslant (1 + \gamma^2),$$

whilst

$$\sum_{j=1}^{\infty}(1+\gamma^2)|\langle u^\dagger, x_j\rangle_X|^2 = (1+\gamma^2)\|u^\dagger\|^2 < +\infty.$$

Therefore, by the dominated convergence theorem, we have

$$\begin{aligned}\lim_{\alpha\to 0}\|R_\alpha f - A^\dagger f\|_X^2 &= \lim_{\alpha\to 0}\sum_{j=1}^{\infty}(\sigma_j g_\alpha(\sigma_j) - 1)^2|\langle u^\dagger, x_j\rangle_X|^2 \\ &= \sum_{j=1}^{\infty}(\lim_{\alpha\to 0}\sigma_j g_\alpha(\sigma_j) - 1)^2|\langle u^\dagger, x_j\rangle_X|^2 = 0,\end{aligned}$$

where the last equality is due to the pointwise convergence of $g_\alpha(\sigma_j)$ to $1/\sigma_j$. Hence, we have $\|R_\alpha f - A^\dagger f\|_X \to 0$ for $\alpha \to 0$ for all $f \in \mathrm{dom}(A^\dagger)$. □

Theorem 3.14. *Let the assumptions of Theorem 3.13 hold and let $\alpha = \alpha(\delta)$ be an a-priori parameter choice rule. Then $(\{R_\alpha\}_\alpha, \alpha)$ with R_α as defined in* (3.5) *is a convergent regularisation method if* $\lim_{\delta\to 0}\alpha(\delta) = 0$ *and*

$$\lim_{\delta\to 0}\delta C_{\alpha(\delta)} = 0.$$

Proof. The result follows immediately from $\|R_{\alpha(\delta)}\|_{\mathcal{L}(X,Y)} \leqslant C_{\alpha(\delta)}$ and Theorem 3.10. □

3.3.1 Truncated singular value decomposition

As a first example of a spectral regularisation of the form (3.5), let us consider the so-called *truncated singular value decomposition*. The idea is to discard all singular values below a certain threshold α, which is achieved using the following function g_α

$$g_\alpha(\sigma) = \begin{cases}\frac{1}{\sigma} & \sigma \geqslant \alpha, \\ 0 & \sigma < \alpha.\end{cases} \tag{3.8}$$

Note that for all $\sigma > 0$ we naturally obtain $\lim_{\alpha\to 0} g_\alpha(\sigma) = 1/\sigma$. Condition (3.7) is obviously satisfied with $\gamma = 1$ and condition (3.6) with $C_\alpha = \frac{1}{\alpha}$. Therefore, truncated SVD is a convergent regularisation if

$$\lim_{\delta\to 0} \frac{\delta}{\alpha} = 0. \tag{3.9}$$

Equation (3.5) then reads as follows

$$R_\alpha f = \sum_{\sigma_j \geqslant \alpha} \frac{1}{\sigma_j} \langle f, y_j \rangle_{\mathcal{Y}} x_j, \tag{3.10}$$

for all $f \in \mathcal{Y}$. Note that the sum in (3.10) is always well-defined (i.e. finite) for any $\alpha > 0$ as zero is the only accumulation point of singular values of compact operators.

Let $A \in \mathcal{K}(\mathcal{X}, \mathcal{Y})$ with singular system $\{(\sigma_j, x_j, y_j)\}_{j\in\mathbf{N}}$, and choose for $\delta > 0$ an index function $j^* : \mathbf{R}_+ \to \mathbf{N}$ with $j^*(\delta) \to \infty$ for $\delta \to 0$ and $\lim_{\delta\to 0} \delta/\sigma_{j^*(\delta)} = 0$. We can then choose $\alpha(\delta) = \sigma_{j^*(\delta)}$ as an a-priori parameter choice rule to obtain a convergent regularisation. Note that in practice a larger δ implies that more and more singular values have to be cut off in order to guarantee a stable recovery that successfully suppresses the data error. A disadvantage of this approach is that it requires the knowledge of the singular vectors of A (only finitely many, but the number can still be large).

3.3.2 Tikhonov regularisation

The main idea behind Tikhonov regularisation[1] is to consider the normal equations and shift the eigenvalues of A^*A by a constant factor, which will be associated with the regularisation parameter α. This shift can be realised via the function

$$g_\alpha(\sigma) = \frac{\sigma}{\sigma^2 + \alpha} \tag{3.11}$$

and the corresponding Tikhonov regularisation (3.5) reads as follows

$$R_\alpha f = \sum_{j=1}^{\infty} \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle f, y_j \rangle_{\mathcal{Y}} x_j. \tag{3.12}$$

Again, we immediately observe that for all $\sigma > 0$ we have $\lim_{\alpha\to 0} g_\alpha(\sigma) = 1/\sigma$. Condition (3.7) is satisfied with $\gamma = 1$. Since $0 \leqslant (\sigma - \sqrt{\alpha})^2 = \sigma^2 - 2\sigma\sqrt{\alpha} + \alpha$, we get that $\sigma^2 + \alpha \geqslant 2\sigma\sqrt{\alpha}$ and

$$\frac{\sigma}{\sigma^2 + \alpha} \leqslant \frac{1}{2\sqrt{\alpha}}.$$

This estimate implies that (3.6) holds with $C_\alpha = \frac{1}{2\sqrt{\alpha}}$. Therefore, Tikhonov regularisation is a convergent regularisation if

$$\lim_{\delta\to 0} \frac{\delta}{\sqrt{\alpha}} = 0. \tag{3.13}$$

The formula (3.12) suggests that we need the full singular system of A in order to compute the regularisation. However, we note that σ_j^2 are the eigenvalues of A^*A and, hence, $\sigma_j^2 + \alpha$ are the

[1] Named after the Russian mathematician Andrey Nikolayevich Tikhonov (30 October 1906 - 7 October 1993)

eigenvalues of $A^*A+\alpha I$ (where I is the identity operator). Applying this operator to the regularised solution $u_\alpha = R_\alpha f$, we get

$$(A^*A + \alpha I)u_\alpha = \sum_{j=1}^{\infty}(\sigma_j^2 + \alpha)\langle u_\alpha, x_j\rangle_X x_j = \sum_{j=1}^{\infty}(\sigma_j^2 + \alpha)\frac{\sigma_j}{\sigma_j^2 + \alpha}\langle f, y_j\rangle_Y x_j = A^*f.$$

Therefore, the regularised solution u_α can be computed without knowing the singular system of A by solving the following well-posed linear equation

$$(A^*A + \alpha I)u_\alpha = A^*f. \tag{3.14}$$

Remark 3.15. Rewriting equation (3.14) as

$$A^*(Au_\alpha - f) + \alpha u_\alpha = 0,$$

we note that it looks like a condition for the minimum of some quadratic form. Indeed, it can be easily checked that (3.14) is the first order optimality condition for the following optimisation problem

$$\min_{u\in X}\frac{1}{2}\|Au - f\|^2 + \alpha\|u\|^2. \tag{3.15}$$

The condition (3.14) is necessary (and, by convexity, sufficient) for the minimum of the functional in (3.15). Therefore, the regularised solution u_α can also be computed by solving (numerically) the variational problem (3.15). This is the starting point for modern variational regularisation methods, which we will consider in the next chapter.

Chapter 4

Variational Regularisation

Recall the variational formulation of Tikhonov regularisation for some data $f_\delta \in \mathcal{Y}$

$$\min_{u \in \mathcal{X}} \|Au - f_\delta\|^2 + \alpha \|u\|^2.$$

The first term in this expression, $\|Au - f_\delta\|^2$, penalises the misfit between the predictions of the operator A and the measured data f_δ and is called the *fidelity function* or *fidelity term*. The second term, $\|u\|^2$ penalises some unwanted features of the solution (in this case, a large norm) and is called the *regularisation term*. The regularisation parameter α in this context balances the influence of these two terms on the functional to be minimised. More generally, using the notation $\mathcal{J}(u)$ for the regulariser, we can formally write down the variational regularisation problem as follows

$$\min_{u \in \mathcal{X}} \frac{1}{2} \|Au - f_\delta\|^2 + \alpha \mathcal{J}(u), \tag{4.1}$$

(the $\frac{1}{2}$ in front of the fidelity term is there to simplify notation later). The regularisation operator R_α is defined as follows

$$R_\alpha f_\delta \in \arg\min_{u \in \mathcal{X}} \frac{1}{2} \|Au - f_\delta\|^2 + \alpha \mathcal{J}(u).$$

In general, the minimiser doesn't have to unique, hence the inclusion and not equality. Other fidelity terms (not just $\|Au - f_\delta\|^2$) are possible and useful in many situations. In this course, however, we will use the squared norm for the sake of simplicity. In this chapter, we will study the properties of (4.1) for different choices of $\mathcal{J}$, but before that we will recall some necessary theoretical concepts.

4.1 Background

4.1.1 Banach spaces and weak convergence

Banach spaces are complete, normed vector spaces (just as Hilbert spaces are), but with a norm that need not be induced by an inner product. For every Banach space $\mathcal{X}$, we can define the associated space of linear and continuous functionals which is called the *dual space* $\mathcal{X}^*$ of $\mathcal{X}$, i.e. $\mathcal{X}^* := \mathcal{L}(\mathcal{X}, \mathbb{R})$. If $u \in \mathcal{X}$ and $p \in \mathcal{X}^*$, then we usually write the *dual product* $\langle p, u \rangle$ instead of $p(u)$. Moreover, for any $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ there exists a unique operator $A^* \colon \mathcal{Y}^* \to \mathcal{X}^*$, called the *adjoint* of A, such that for all $u \in \mathcal{X}$ and $p \in \mathcal{Y}^*$ we have

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

It is easy to see that either side of the equation are well-defined, e.g. $A^*p \in \mathcal{X}^*$ and $u \in \mathcal{X}$. The dual space of a Banach space $\mathcal{X}$ can be equipped with the following norm

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

With this norm the dual space is itself a Banach space. Therefore, it has a dual space as well which we will call the bi-dual space of $\mathcal{X}$ and denote it with $\mathcal{X}^{**} := (\mathcal{X}^*)^*$. As every $u \in \mathcal{X}$ defines a continuous and linear mapping on the dual space $\mathcal{X}^*$ by

$$\langle E(u), p\rangle := \langle p, u\rangle,$$

the mapping $E\colon \mathcal{X} \to \mathcal{X}^{**}$ is well-defined. It can be shown that E is a linear and continuous isometry (and thus injective). In the special case when E is surjective, we call $\mathcal{X}$ *reflexive*. Examples of reflexive Banach spaces include Hilbert spaces and L^p, ℓ^p spaces with $1 < p < \infty$. We call the space $\mathcal{X}$ *separable* if there exists a set $\mathcal{X}' \subset \mathcal{X}$ of at most countable cardinality such that $\overline{\mathcal{X}'} = \mathcal{X}$.

A problem in infinite-dimensional spaces is that bounded sequences may fail to have convergent subsequences. An example in ℓ^2 is the sequence $\{u^k\}_{k\in\mathbf{N}} \subset \ell^2, u^k_j = 1$ if $k = j$ and 0 otherwise. It is easy to see that $\|u^k\|_{\ell^2} = 1$ and that there is no $u \in \ell^2$ such that $u^k \to u$. To circumvent this problem, we define a weaker topology on $\mathcal{X}$. We say that $\{u^k\}_{k\in\mathbf{N}} \subset \mathcal{X}$ *converges weakly* to $u \in \mathcal{X}$ if and only if for all $p \in \mathcal{X}^*$ the sequence of real numbers $\{\langle p, u^k\rangle\}_{k\in\mathbf{N}}$ converges and

$$\langle p, u_j\rangle \to \langle p, u\rangle.$$

We will denote weak convergence by $u^k \rightharpoonup u$. On a dual space $\mathcal{X}^*$ we could define another topology (in addition to the strong topology induced by the norm and the weak topology as the dual space is a Banach space as well). We say a sequence $\{p^k\}_{k\in\mathbf{N}} \subset \mathcal{X}^*$ *converges weakly-** to $p \in \mathcal{X}^*$ if and only if

$$\langle p^k, u\rangle \to \langle p, u\rangle \quad \text{for all } u \in \mathcal{X}$$

and we denote weak-$*$ convergence by $p^k \rightharpoonup^* p$. Similarly, for any topology τ on $\mathcal{X}$ we denote the convergence in that topology by $u^k \xrightarrow{\tau} u$. with these two new notions of convergence, we can solve the problem of bounded sequences not necessarily having convergent subsequences:

Theorem 4.1 (Banach-Alaoglu Theorem, e.g. [17, p. 70] or [21, p. 141])**.** *Let $\mathcal{X} = (\mathcal{X}^\diamond)^*$ be the dual of a Banach space $\mathcal{X}^\diamond$. Then the unit ball $\mathcal{B}_{\mathcal{X}} = \{u \in \mathcal{X} | \|x\| \leqslant 1\}$ is compact in the weak-* topology. If $\mathcal{X}^\diamond$ is separable, then the weak-* topology is metrisable on bounded sets and every bounded sequence $\{u^k\}_{k\in\mathbf{N}} \subset \mathcal{X}$ has a weak-* convergent subsequence.*

Theorem 4.2 ([23, p. 64])**.** *Each bounded sequence $\{u^k\}_{k\in\mathbf{N}}$ in a reflexive Banach space $\mathcal{X}$ has a weakly convergent subsequence.*

An important property of functionals, which we will need later, is sequential lower semicontinuity. Roughly speaking this means that the functional values for arguments near an argument u are either close to $E(u)$ or greater than $E(u)$.

Definition 4.3. *Let $\mathcal{X}$ be a Banach space with topology $\tau_{\mathcal{X}}$. The functional $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ is said to be* sequentially lower semi-continuous with respect to $\tau_{\mathcal{X}}$ *($\tau_{\mathcal{X}}$-l.s.c.) at $u \in \mathcal{X}$ if*

$$E(u) \leqslant \liminf_{j\to\infty} E(u_j)$$

for all sequences $\{u_j\}_{j\in\mathbf{N}} \subset \mathcal{X}$ with $u_j \to u$ in the topology $\tau_{\mathcal{X}}$ of $\mathcal{X}$.

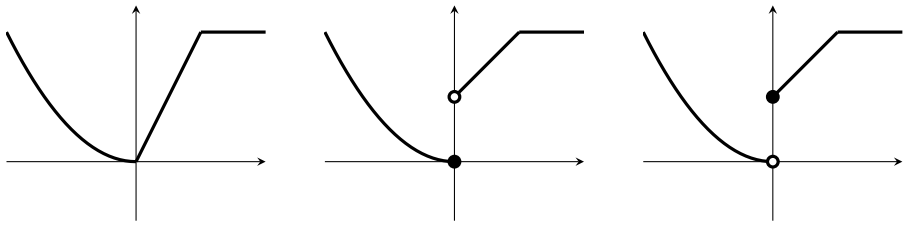

Figure 4.1: Visualisation of lower semi-continuity. The solid dot at a jump indicates the value that the function takes. The function on the left is continuous and thus lower semi-continuous. The functions in the middle and on the right are discontinuous. While the function in the middle is lower semi-continuous, the function on the right is not (due to the limit from the left at the discontinuity).

Remark 4.4. For topologies that are not induced by a metric we have to distinguish between a topological property and its sequential version, e.g. continuity and sequential continuity. If the topology is induced by a metric, these concepts are equivalent. However, this is not generally the case: for instance, the weak and weak-$*$ topology are generally not induced by a metric.

Example 4.5. The functional $\|\cdot\|_1 : \ell^2 \to \bar{\mathbf{R}}$ with

$$\|u\|_1 = \begin{cases} \sum_{j=1}^{\infty} |u_j| & \text{if } u \in \ell^1, \\ \infty & \text{else,} \end{cases}$$

is weakly (and, hence, strongly) lower semi-continuous in ℓ^2.

Proof. Let $\{u^j\}_{j\in\mathbf{N}} \subset \ell^2$ be a weakly convergent sequence with $u^j \rightharpoonup u \in \ell^2$. We have with $\delta_k : \ell^2 \to \mathbf{R}, \langle \delta_k, v\rangle = v_k$ that for all $k \in \mathbf{N}$

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

The assertion follows then with Fatou's lemma

$$\|u\|_1 = \sum_{k=1}^{\infty} |u_k| = \sum_{k=1}^{\infty} \lim_{j\to\infty} |u_k^j| \leqslant \liminf_{j\to\infty} \sum_{k=1}^{\infty} |u_k^j| = \liminf_{j\to\infty} \|u^j\|_1.$$

Note that neither the left hand side nor the right hand side need to be finite. □

Infinity calculus

We will look at functionals $E : \mathcal{X} \to \bar{\mathbf{R}}$ whose range is modelled to be the *extended real line* $\bar{\mathbf{R}} := \mathbf{R} \cup \{-\infty, +\infty\}$ where the symbol $+\infty$ denotes an element that is not part of the real line that is by definition larger than any other element of the reals, i.e.

$$x < +\infty$$

for all $x \in \mathbf{R}$ (similarly, $x > -\infty$ for all $x \in \mathbf{R}$). This is useful to model constraints: for instance, if we were trying to minimise $E : [-1, \infty) \to \mathbf{R}, x \mapsto x^2$ we could remodel this minimisation problem by $\widetilde{E} : \mathbf{R} \to \bar{\mathbf{R}}$

$$\widetilde{E}(x) = \begin{cases} x^2 & \text{if } x \geqslant -1, \\ \infty & \text{else.} \end{cases}$$

Obviously both functionals have the same minimiser but $\widetilde{E}$ is defined on a vector space and not only on a subset. This has two important consequences: on the one hand, it makes many theoretical arguments easier as we do not need to worry whether $E(x + y)$ is defined or not. On the other hand, it makes practical implementations easier as we are dealing with unconstrained optimisation instead of constrained optimisation. This comes at a cost that some algorithms are not applicable any more, e.g. the function $\widetilde{E}$ is not differentiable everywhere whereas E is (in the interior of its domain).

It is useful to note that one can calculate on the extended real line $\bar{\mathbf{R}}$ as we are used to on the real line $\mathbf{R}$ but the operations with $\pm\infty$ need yet to be defined.

Definition 4.6. *The extended real line is defined as* $\bar{\mathbf{R}} := \mathbf{R} \cup \{-\infty, +\infty\}$ *with the following rules that hold for any* $x \in \mathbf{R}$ *and* $\lambda > 0$*:*

$$\begin{aligned} x \pm \infty := \pm\infty + x &:= \pm\infty \\ \lambda \cdot (\pm\infty) := \pm\infty \cdot \lambda &:= \pm\infty, \quad -1 \cdot (\pm\infty) := \mp\infty \\ x/(\pm\infty) &:= 0 \\ \infty + \infty := \infty, \quad -\infty - \infty &:= -\infty. \end{aligned}$$

Some calculations are *not defined*, e.g.,

$$+\infty - \infty \text{ and } \quad (\pm\infty) \cdot (\pm\infty).$$

Using functions with values on the extended real line, one can easily describe sets $C \subset \mathcal{X}$.

Definition 4.7 (Characteristic function)**.** *Let* $C \subset \mathcal{X}$ *be a set. The function* $\chi_C \colon \mathcal{X} \to \bar{\mathbf{R}}$*,*

$$\chi_C(u) = \begin{cases} 0 & u \in C, \\ \infty & u \in \mathcal{X} \setminus C, \end{cases}$$

is called the characteristic function of the set C*.*

Using characteristic functions, one can easily write constrained optimisation problems as unconstrained ones:

$$\min_{u \in C} E(u) \quad \Leftrightarrow \quad \min_{u \in \mathcal{X}} E(u) + \chi_C(u).$$

Definition 4.8. *Let* $\mathcal{X}$ *be a vector space and* $E \colon \mathcal{X} \to \bar{\mathbf{R}}$ *a functional. Then the* effective domain *of* E *is*

$$\operatorname{dom}(E) := \{u \in \mathcal{X} \mid E(u) < \infty\}.$$

Definition 4.9. *A functional* E *is called* proper *if the effective domain* $\operatorname{dom}(E)$ *is not empty and* $E(u) \neq -\infty$ *for all* $u \in \mathcal{X}$*.*

4.1.2 Existence of minimisers

Definition 4.10. *Let* $E \colon \mathcal{X} \to \bar{\mathbf{R}}$ *be a functional. We say that* $u^* \in \mathcal{X}$ *solves the minimisation problem*

$$\min_{u \in \mathcal{X}} E(u)$$

if and only if $E(u^*) \neq \pm\infty$ *and* $E(u^*) \leqslant E(u)$*, for all* $u \in \mathcal{X}$*. We call* u^* *a* minimiser *of* E*.*

Definition 4.11. *A functional* $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ *is called* bounded from below *if there exists a constant* $C > -\infty$ *such that for all* $u \in \mathcal{X}$ *we have* $E(u) \geqslant C$.

This condition is obviously necessary for the finiteness of the infimum $\inf_{u\in\mathcal{X}} E(u)$.

If all minimising sequences (that converge to the infimum assuming it exists) are unbounded, then there cannot exist a minimiser. A sufficient condition to avoid such a scenario is *coercivity*.

Definition 4.12. *A functional* $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ *is called* coercive, *if for any* $C \subseteq \mathcal{X}$, $\sup_{u\in C} E(u) < \infty$ *implies that* $\sup_{u\in C} \|u\| < \infty$.

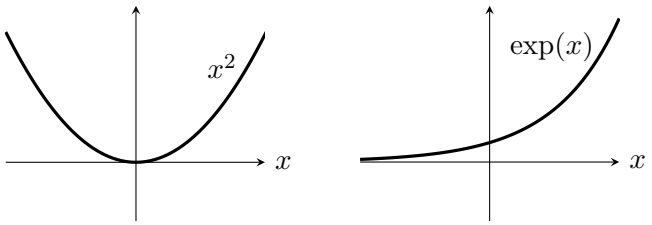

Figure 4.2: While the coercive function on the left has a minimiser, it is easy to see that the non-coercive function on the right does not have a minimiser.

Remark 4.13. This definition of coercivity can equivalently be reformulated using its contrapositive: for any $C \subseteq \mathcal{X}$, if $\sup_{u\in C} \|u\| = \infty$, then $\sup_{u\in C} E(u) = \infty$. Another useful way to think of coercivity is in terms of the sub-level sets: E is coercive if and only if $\{u \in \mathcal{X} | E(u) \leqslant \alpha\}$ is a bounded set for any $\alpha \in \mathbf{R}$.

Although coercivity is not strictly speaking necessary, it is sufficient to ensure that all minimising sequences are bounded.

Lemma 4.14. *Let* $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ *be a proper, coercive functional which is bounded from below. Then the infimum* $\inf_{u\in\mathcal{X}} E(u)$ *exists in* $\mathbf{R}$, *there are minimising sequences, i.e.* $\{u_j\}_{j\in\mathbf{N}} \subset \mathcal{X}$ *with* $E(u_j) \to \inf_{u\in\mathcal{X}} E(u)$, *and all minimising sequences are bounded.*

Proof. As E is proper and bounded from below, there exists a $C_1 > 0$ such that we have $-\infty < -C_1 < \inf_u E(u) < \infty$ which also guarantees the existence of a minimising sequence. Let $\{u_j\}_{j\in\mathbf{N}}$ be any minimising sequence, i.e. $E(u_j) \to \inf_u E(u)$. Then there exists a $j_0 \in \mathbf{N}$ such that for all $j > j_0$ we have

$$E(u_j) \leqslant \underbrace{\inf_u E(u) + 1}_{=:C_2} < \infty.$$

With $C := \max\{C_1, C_2\}$ we have that $|E(u_j)| < C$ for all $j > j_0$ and thus from the coercivity it follows that $\{u_j\}_{j>j_0}$ is bounded, see Remark 4.13. Including a finite number of elements does not change its boundedness which proves the assertion. □

Remark 4.1. Note that the proof above does not change materially if E is not bounded from below: in that case, there are still "minimising sequences" $\{u_j\}_{j\in\mathbf{N}}$ (meaning that $E(u_j) \to -\infty$), and the same argument shows that these sequences must be bounded. Note, however, that in the following we will require boundedness from below, since we have included in our definition of minimisers (Definition 4.10) that the value of the functional at the minimiser is real.

A positive answer about the existence of minimisers is given by the following Theorem known as the "direct method" or "fundamental theorem of optimisation".

Theorem 4.15 ("Direct method", David Hilbert, around 1900)**.** *Let $\mathcal{X}$ be a Banach space and $\tau_{\mathcal{X}}$ a topology (not necessarily the one induced by the norm) on $\mathcal{X}$ such that bounded sequences have $\tau_{\mathcal{X}}$-convergent subsequences. Let $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ be proper, bounded from below, coercive and $\tau_{\mathcal{X}}$-l.s.c. Then E has a minimiser.*

Proof. From Lemma 4.14 we know that $\inf_{u\in\mathcal{X}} E(u)$ is finite, minimising sequences exist and that they are bounded. Let $\{u_j\}_{j\in\mathbf{N}} \in \mathcal{X}$ be a minimising sequence. Thus, from the assumption on the topology $\tau_{\mathcal{X}}$ there exists a subsequence $\{u_{j_k}\}_{k\in\mathbf{N}}$ and $u^* \in \mathcal{X}$ with $u_{j_k} \overset{\tau_{\mathcal{X}}}{\to} u^*$ for $k \to \infty$. From the sequential lower semi-continuity of E we obtain

$$-\infty < \inf_{u\in\mathcal{X}} E(u) \leqslant E(u^*) \leqslant \liminf_{k\to\infty} E(u_{j_k}) = \lim_{j\to\infty} E(u_j) = \inf_{u\in\mathcal{X}} E(u) < \infty,$$

which shows that $E(u^*) \neq \pm\infty$ and $E(u^*) \leqslant E(u)$ for all $u \in \mathcal{X}$; thus u^* is a minimiser of E. □

The above theorem is very general, and its conditions may be hard to verify, but the situation is a bit easier in *reflexive* Banach spaces (thus also in Hilbert spaces).

Corollary 4.16. Let $\mathcal{X}$ be a reflexive Banach space and $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ be a functional which is proper, bounded from below, coercive and l.s.c. with respect to the weak topology. Then there exists a minimiser of E.

Proof. The statement follows from the direct method, Theorem 4.15, as in reflexive Banach spaces bounded sequences have weakly convergent subsequences, see Theorem 4.2. □

Remark 4.17. For convex functionals, the situation is even easier. It can be shown that a convex function is l.s.c. with respect to the weak topology if and only if it is l.s.c. with respect to the strong topology (see e.g. [11, Corollary 2.2]).

Remark 4.18. It is easy to see that the key ingredient for the existence of minimisers is that bounded sequences have a convergent subsequence. In variational regularisation this is usually ensured by an appropriate choice of the regularisation functional.

4.1.3 Convex analysis

A property of fundamental importance of sets and functions is convexity.

Definition 4.19. *Let $\mathcal{X}$ be a vector space. A subset $C \subset \mathcal{X}$ is called* convex*, if $\lambda u + (1-\lambda)v \in C$ for all $\lambda \in (0,1)$ and all $u, v \in C$.*

Definition 4.20. *A functional $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ is called* convex*, if*

$$E(\lambda u + (1-\lambda)v) \leqslant \lambda E(u) + (1-\lambda)E(v)$$

for all $\lambda \in (0,1)$ and all $u, v \in \operatorname{dom}(E)$ with $u \neq v$. It is called strictly convex *if the inequality is strict. It is called* strongly *convex with constant $\theta > 0$ if $E(u) - \theta\|u\|^2$ is convex.*

Obviously, strong convexity implies strict convexity and strict convexity implies convexity.

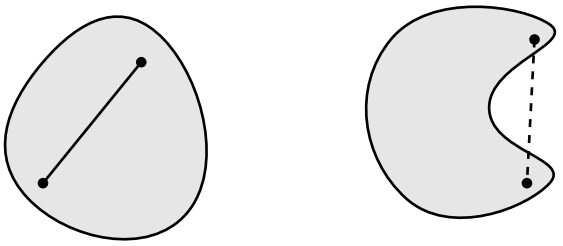

Figure 4.3: Example of a convex set (left) and non-convex set (right).

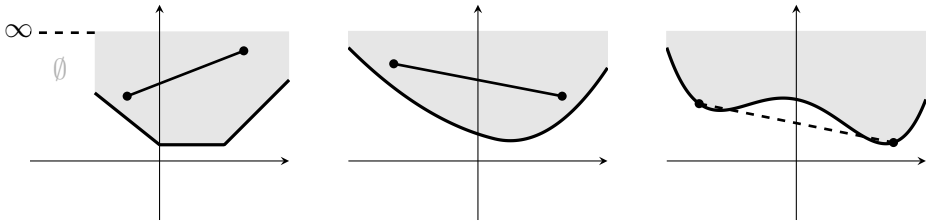

Figure 4.4: Example of a convex function (left), a strictly convex function (middle) and a non-convex function (right).

Example 4.21. The absolute value function $\mathbf{R} \to \mathbf{R}, x \mapsto |x|$ is convex but not strictly convex. The quadratic function $x \mapsto x^2$ is strongly (and hence strictly) convex. The function $x \mapsto x^4$ is strictly convex, but not strongly convex. For other examples, see Figure 4.4.

Example 4.22. The characteristic function $\chi_C(u)$ is convex if and only if C is a convex set. To see the convexity, let $u, v \in \mathrm{dom}(\chi_C) = C$. Then by the convexity of C the convex combination $\lambda u + (1-\lambda)v$ is as well in C and both the left and the right hand side of the desired inequality are zero.

Lemma 4.23. *Let $\alpha \geqslant 0$ and $E, F \colon \mathcal{X} \to \bar{\mathbf{R}}$ be two convex functionals. Then $E + \alpha F \colon \mathcal{X} \to \bar{\mathbf{R}}$ is convex. Furthermore, if $\alpha > 0$ and F strictly convex, then $E + \alpha F$ is strictly convex.*

Fenchel conjugate

In convex optimisation problems (i.e. those involving convex functions) the concept of *Fenchel conjugates* plays a very important role.

Definition 4.24. *Let $E \colon \mathcal{X} \to \bar{\mathbf{R}}$ be a functional. The functional $E^* \colon \mathcal{X}^* \to \bar{\mathbf{R}}$,*

$$E^*(p) = \sup_{u \in \mathcal{X}} [\langle p, u \rangle - E(u)],$$

is called the Fenchel conjugate of E.

Theorem 4.25 ([11, Proposition 4.1])**.** *For any functional $E \colon \mathcal{X} \to \bar{\mathbf{R}}$ the following inequality holds:*

$$E^{**} := (E^*)^* \leqslant E.$$

If E is proper, lower-semicontinuous (see Def. 4.3) and convex, then

$$E^{**} = E.$$

Note that E^{**} is in principle defined on $\mathcal{X}^{**}$, but we interpret it as a functional on $\mathcal{X}$, understood as a subspace of $\mathcal{X}^{**}$.

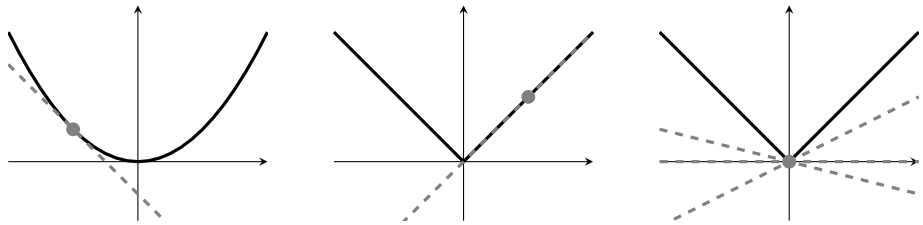

Figure 4.5: Visualisation of the subdifferential. Linear approximations of the functional have to lie completely underneath the function. For points where the function is not differentiable there may be more than one such approximation.

Subgradients

For convex functions, one can generalise the concept of a derivative so that it also makes sense for non-differentiable functions.

Definition 4.26. *A functional* $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ *is called* subdifferentiable *at* $u \in \mathcal{X}$*, if there exists an element* $p \in \mathcal{X}^*$ *such that*

$$E(v) \geqslant E(u) + \langle p, v - u \rangle$$

holds, for all $v \in \mathcal{X}$*. Furthermore, we call* p *a* subgradient *at position* u*. The collection of all subgradients at position* u*, i.e.*

$$\partial E(u) := \{p \in \mathcal{X}^* \mid E(v) \geqslant E(u) + \langle p, v - u \rangle \, , \forall v \in \mathcal{X}\} \, ,$$

is called subdifferential *of* E *at* u*.*

It is clear that if a convex functional $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ is proper, i.e. $\operatorname{dom}(E) \neq \emptyset$, then for all $u \notin \operatorname{dom}(E)$ the subdifferential is empty. A sufficient (but not necessary) condition for E to have a subgradient at $u \in \operatorname{dom}(E)$ is given by

Proposition 4.27 ([11, Proposition 5.2])**.** *Let* $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ *be a convex functional and* $u \in \operatorname{dom}(E)$ *such that* E *is continuous at* u*. Then* $\partial E(u) \neq \emptyset$*.*

Theorem 4.28 ([3, Theorem 7.13])**.** *Let* $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ *be a proper convex function and* $u \in \operatorname{dom}(E)$*. Then* $\partial E(u)$ *is a weak-*$*$ *compact convex subset of* $\mathcal{X}^*$*.*

For differentiable functions, the subdifferential consists of just one element: the derivative. For non-differentiable functionals, the subdifferential is set-valued; let us consider the subdifferential of the absolute value function as an illustrative example.

Example 4.29. Let $E\colon \mathbf{R} \to \mathbf{R}$ be the absolute value function $E(u) = |u|$. Then, the subdifferential of E at u is given by

$$\partial E(u) = \begin{cases} \{1\} & \text{for } u > 0, \\ [-1, 1] & \text{for } u = 0, \\ \{-1\} & \text{for } u < 0, \end{cases}$$

which you will prove as an exercise. A visual explanation is given in Figure 4.5.

The subdifferential of a sum of two functions can be characterised as follows.

Theorem 4.30 ([11, Proposition 5.6])**.** *Let $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ and $F\colon \mathcal{X} \to \bar{\mathbf{R}}$ be proper l.s.c. convex functions and suppose $\exists u \in \operatorname{dom}(E) \cap \operatorname{dom}(F)$ such that E is continuous at u. Then*

$$\partial(E + F) = \partial E + \partial F.$$

Using the subdifferential, one can characterise minimisers of convex functionals.

Theorem 4.31. *An element $u \in \mathcal{X}$ is a minimiser of the functional $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ if and only if $0 \in \partial E(u)$.*

Proof. By definition, $0 \in \partial E(u)$ if and only if for all $v \in \mathcal{X}$ it holds

$$E(v) \geqslant E(u) + \langle 0, v - u\rangle = E(u),$$

which is by definition the case if and only if u is a minimiser of E. □

The next result connects subgradients and convex conjugates

Theorem 4.32 ([11, Proposition 5.1])**.** *Let $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ be a convex function and $E^*\colon \mathcal{X}^* \to \bar{\mathbf{R}}$ its convex conjugate. Then $p \in \partial E(u)$ if and only if*

$$E(u) + E^*(p) = \langle p, u\rangle.$$

Proof. Left as an exercise. □

Bregman divergences

Convex functions naturally define some distance measure that became known as the Bregman divergence.

Definition 4.33. *Let $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ be a convex functional. Moreover, let $u, v \in \mathcal{X}, E(v) < \infty$ and $q \in \partial E(v)$. Then the* (generalised) Bregman divergence *of E between u and v is defined as*

$$D_E^q(u, v) := E(u) - E(v) - \langle q, u - v\rangle. \tag{4.2}$$

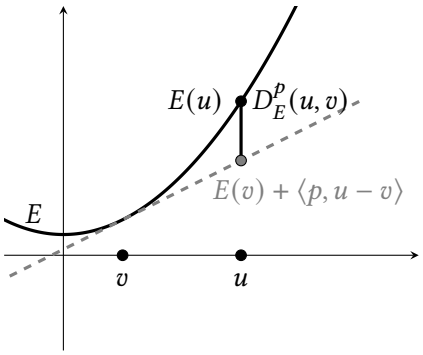

Figure 4.6: Visualization of the Bregman divergence.

Remark 4.34. It is easy to check that a Bregman divergence somewhat resembles a metric as for all $u, v \in \mathcal{X}, q \in \partial E(v)$ we have that $D_E^q(u, v) \geqslant 0$ and $D_E^q(v, v) = 0$. There are functionals where the Bregman divergence (up to a square root) is actually a metric; e.g. if $E(u) := \frac{1}{2}\|u\|_{\mathcal{X}}^2$ in a Hilbert space $\mathcal{X}$, then $D_E^q(u, v) = \frac{1}{2}\|u - v\|_{\mathcal{X}}^2$. However, in general, Bregman divergences are not symmetric and $D_E^q(u, v) = 0$ does not imply $u = v$, as you will see on the examples sheets.

To overcome the issue of non-symmetry, one can introduce the so-called *symmetric Bregman divergence.*

Definition 4.35. *Let $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ be a convex functional. Moreover, let $u, v \in \mathcal{X}, E(u) < \infty, E(v) < \infty$, $q \in \partial E(v)$ and $p \in \partial E(u)$. Then the* symmetric Bregman divergence *of E between u and v is defined as*

$$D_E^{\text{symm}}(u, v) := D_E^q(u, v) + D_E^p(v, u) = \langle p - q, u - v \rangle. \tag{4.3}$$

Absolutely one-homogeneous functionals

Definition 4.36. *A functional $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ is called absolutely one-homogeneous if*

$$E(\lambda u) = |\lambda| E(u) \quad \forall \lambda \in \mathbf{R}, \quad \forall u \in \mathcal{X}.$$

Absolutely one-homogeneous convex functionals have some useful properties, for example, it is obvious that $E(0) = 0$. Some further properties are listed below.

Proposition 4.37. *Let $E(\cdot)$ be a convex absolutely one-homogeneous functional and let $p \in \partial E(u)$. Then the following equality holds:*

$$E(u) = \langle p, u \rangle.$$

Proof. Left as exercise. □

Remark 4.38. The Bregman divergence $D_E^p(v, u)$ in this case can be written as follows:

$$D_E^p(v, u) = E(v) - \langle p, v \rangle.$$

Proposition 4.39. *Let $E(\cdot)$ be a proper, convex, l.s.c. and absolutely one-homogeneous functional. Then the Fenchel conjugate $E^*(\cdot)$ is the characteristic function of the convex set $\partial E(0)$.*

Proof. Left as exercise. □

An obvious consequence of the above results is the following

Proposition 4.40. *For any $u \in \mathcal{X}$, $p \in \partial E(u)$ if and only if $p \in \partial E(0)$ and $E(u) = \langle p, u \rangle$.*

Uniqueness of minimisers

Theorem 4.41. *Assume that the functional $E\colon \mathcal{X} \to \bar{\mathbf{R}}$ has at least one minimiser and is strictly convex. Then the minimiser is unique.*

Proof. Let u, v be two minimisers of E and assume that they are different, i.e. $u \neq v$. Then it follows from the minimising properties of u and v as well as the strict convexity of E that

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

which is a contradiction. Thus, $u = v$ and the assertion is proven. □

Example 4.42. Convex (but not strictly convex) functions may have have more than one minimiser, examples include constant and trapezoidal functions, see Figure 4.7. On the other hand, convex (and even non-convex) functions may have a unique minimiser, see Figure 4.7.

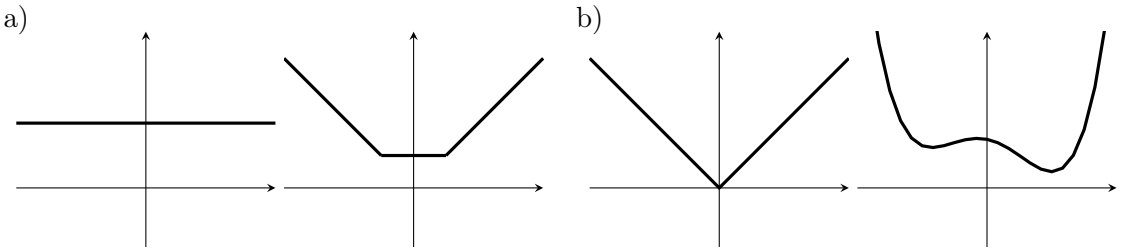

Figure 4.7: a) Convex functions may not have a unique minimiser. b) Neither strict convexity nor convexity is necessary for the uniqueness of a minimiser.

4.2 Well-posedness and Regularisation Properties

Our goal is to study the properties of optimisation problem (4.1) as a convergent regularisation for the ill-posed problem

$$Au = f, \tag{4.4}$$

where $A\colon \mathcal{X} \to \mathcal{Y}$ is a linear bounded operator, $\mathcal{Y}$ is a Banach space and $\mathcal{X}$ is the dual of a separable Banach space. In particular, we will ask questions of existence of minimisers (well-posedness of the regularised problem) and parameter choice rules that guarantee the convergence of the minimisers to an appropriate generalised solution of (4.4) for different choices of the regularisation functional. To this end, we need to extend the definition of a minimum-norm solution (Definition 2.3) to an arbitrary regularisation term.

Definition 4.43 ($\mathcal{J}$-minimising solutions)**.** *Let $u^\dagger_{\mathcal{J}}$ be a least-squares solution, i.e.*

$$\|Au^\dagger_{\mathcal{J}} - f\|_{\mathcal{Y}} = \inf\{\|Av - f\|_{\mathcal{Y}} | v \in \mathcal{X}\}$$

and

$$\mathcal{J}(u^\dagger_{\mathcal{J}}) \leqslant \mathcal{J}(\widetilde{u}) \quad \text{for all least-squares solutions } \widetilde{u}.$$

Then $u^\dagger_{\mathcal{J}}$ is called a $\mathcal{J}$-minimising solution of (4.4).

We will assume that there exists a least-squares solution with a finite value of $\mathcal{J}$, i.e. there exists at least one element u such that $\|Au - f\|_{\mathcal{Y}} = \inf\{\|Av - f\|_{\mathcal{Y}},\ v \in \mathcal{X}\}$ and $\mathcal{J}(u) < +\infty$.

Remark 4.44. A $\mathcal{J}$-minimising solution may not exist and if it does, it may be non-unique. We will later see conditions, under which a $\mathcal{J}$-minimising solution exists. Non-uniqueness, however, is common with popular choices of $\mathcal{J}$. In this case we need to define a *selection operator* that will select a single element from all the $\mathcal{J}$-minimising solutions (see [6]). We will not explicitly mention this, stating all results for just *a* $\mathcal{J}$-minimising solution.

We will need the following result:

Lemma 4.45. *Let $\mathcal{J}(u) = \sum_{i=1}^n \mathcal{J}_i(u)$, where each $\mathcal{J}_i(u)$ is convex and p_i-homogeneous (so that $p_i \geqslant 1$), that is,*

$$\mathcal{J}_i(\lambda u) = |\lambda|^{p_i} \mathcal{J}_i(u) \quad \forall u \in \mathcal{X}, \lambda \in \mathbf{R}.$$

Then the set

$$\ker(\mathcal{J}) := \{u \in \mathcal{X} | \mathcal{J}(u) = 0\}$$

is a linear subspace of $\mathcal{X}$.

Proof. First of all, we note that $\mathcal{J}_i(u) \geqslant 0$ for all $u \in \mathcal{X}$. Indeed, we have

$$0 = \mathcal{J}_i(0) = \mathcal{J}_i\left(\frac{1}{2}u - \frac{1}{2}u\right) \leqslant \frac{1}{2}\mathcal{J}_i(u) + \frac{1}{2}\mathcal{J}_i(-u) = \mathcal{J}_i(u).$$

Now let $u, v \in \ker(\mathcal{J})$ be arbitrary. Then $\mathcal{J}_i(u) = \mathcal{J}_i(v) = 0$ for all $i = 1, ..., n$, hence for any $\lambda \in \mathbf{R}$

$$\begin{aligned} 0 \leqslant \mathcal{J}_i(\lambda u + v) &= 2^{p_i}\mathcal{J}_i\left(\frac{\lambda u}{2} + \frac{v}{2}\right) \leqslant 2^{p_i}\left(\frac{1}{2}\mathcal{J}_i\left(\frac{\lambda u}{2}\right) + \frac{1}{2}\mathcal{J}_i\left(\frac{v}{2}\right)\right) \\ &= \frac{1}{2}\mathcal{J}_i(\lambda u) + \frac{1}{2}\mathcal{J}_i(v) = \frac{|\lambda|^{p_i}}{2}\mathcal{J}_i(u) + \frac{1}{2}\mathcal{J}_i(v) = 0. \end{aligned}$$

Therefore, $\mathcal{J}_i(\lambda u + v) = 0$ for all i and hence $\mathcal{J}(\lambda u + v) = 0$. □

If $\dim\ker(\mathcal{J}) < \infty$, the subspace $\ker(\mathcal{J})$ is *complemented* in $\mathcal{X}$ [3, Theorem 5.89], i.e. there exists a closed subspace $\mathcal{X}_0 \subset \mathcal{X}$ such that $\mathcal{X}_0 \cap \ker(\mathcal{J}) = \{0\}$ and

$$\mathcal{X} = \mathcal{X}_0 \oplus \ker(\mathcal{J}). \tag{4.5}$$

We will use this to establish coercivity of the functional (4.1).

Lemma 4.46. *Suppose that the regularisation functional $\mathcal{J} : \mathcal{X} \to \bar{\mathbf{R}}_+$ is proper, convex and satisfies conditions of Lemma (4.45) and let $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be a bounded linear operator. Suppose also that*

(i) $\dim\ker(\mathcal{J}) < \infty$ and $\mathcal{J}$ is coercive on $\mathcal{X}_0$, where $\mathcal{X}_0$ is such that $\mathcal{X} = \mathcal{X}_0 \oplus \ker(\mathcal{J})$;

(ii) the kernels of A and $\mathcal{J}$ have a trivial intersection, i.e. $\ker(A) \cap \ker(\mathcal{J}) = \{0\}$.

Then the function

$$\Phi_\alpha(u) := \frac{1}{2}\|Au - f\|_{\mathcal{Y}}^2 + \alpha\mathcal{J}(u)$$

is coercive on $\mathcal{X}$ for any $\alpha > 0$.

Proof. Let $\{u_j\}_{j\in\mathbf{N}}$ be a sequence in $\mathcal{X}$. Due to (4.5), there exists a unique decomposition

$$u_j = u_j^0 + u_j^{\ker}, \quad u_j^0 \in \mathcal{X}_0,\ u_j^{\ker} \in \ker(\mathcal{J}).$$

Suppose that $\Phi_\alpha(u_j) \leqslant C$ for all $j \in \mathbf{N}$. Then $\mathcal{J}(u_j) \leqslant C$ and

$$\mathcal{J}(u_j^0) = \mathcal{J}(u_j^0 + u_j^{\ker} - u_j^{\ker}) \leqslant 2^{p^*-1}(\mathcal{J}(u_j) + \mathcal{J}(u_j^{\ker})) = 2^{p^*-1}\mathcal{J}(u_j) \leqslant 2^{p^*-1}C,$$

where $p^* = \max\{p_1, \ldots, p_n\}$. Since $\mathcal{J}$ is coercive on $\mathcal{X}_0$, we get that $\|u_j^0\| \leqslant C'$. Now, define

$$\widetilde{A}\colon \ker(\mathcal{J}) \to A\ker(\mathcal{J}), \quad \widetilde{A} = A|_{\ker(\mathcal{J})}.$$

That is, $\widetilde{A}$ is the restriction of A to $\ker(\mathcal{J})$. Clearly, $\widetilde{A}$ is surjective and by assumption (ii) it is also injective. Since $\ker(\mathcal{J})$ (and, consequently, $A\ker(\mathcal{J})$) is finite-dimensional, $\widetilde{A}^{-1}$ exists and is bounded. Denote $\|\widetilde{A}^{-1}\| =: \widetilde{C}$. Then

$$\begin{aligned} \|u_j^{\ker}\| &= \|\widetilde{A}^{-1}(\widetilde{A}u_j^{\ker})\| \leqslant \widetilde{C}\|Au_j^{\ker}\| = \widetilde{C}\|Au_j^{\ker} + Au_j^0 - f - (Au_j^0 - f)\| \\ &\leqslant \widetilde{C}\Big\|Au_j - f\| + \|Au_j^0 - f\|\Big) \leqslant \widetilde{C}(C + \|A\|\|u_j^0\| + \|f\|) \leqslant C''. \end{aligned}$$

Therefore,

$$\|u_j\| = \|u_j^0 + u_j^{\ker}\| \leqslant \|u_j^0\| + \|u_j^{\ker}\| \leqslant C''',$$

which means that Φ_α is coercive. □

Now we are ready to establish the existence of a $\mathcal{J}$-minimising solution and a regularised solution for any $\alpha > 0$.

Theorem 4.47. *Let $\mathcal{X}$ and $\mathcal{Y}$ be a Banach spaces and $\tau_{\mathcal{X}}$ and $\tau_{\mathcal{Y}}$ some topologies (not necessarily induced by the norm) in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Assume that*

(i) bounded sequences in $\mathcal{X}$ have $\tau_{\mathcal{X}}$-convergent subsequences;

(ii) $\mathcal{J} : \mathcal{X} \to \bar{\mathbf{R}}_+$ is proper, convex $\tau_{\mathcal{X}}$-l.s.c. and satisfies assumptions of Lemma 4.46;

(iii) $A \colon \mathcal{X} \to \mathcal{Y}$ is $\tau_{\mathcal{X}} \to \tau_{\mathcal{Y}}$ continuous;

(iv) $\|\cdot\|_{\mathcal{Y}}$ is $\tau_{\mathcal{Y}}$-lower semicontinuous;

Then

(i') there exists a $\mathcal{J}$-minimising solution $u^{\dagger}_{\mathcal{J}}$ of (4.4);

(ii') for any fixed $\alpha > 0$ and $f \in \mathcal{Y}$ there exists a minimiser

$$u^{\alpha} \in \operatorname*{arg\,min}_{u \in \mathcal{X}} \frac{1}{2}\|Au - f\|^2_{\mathcal{Y}} + \alpha \mathcal{J}(u). \tag{4.6}$$

Proof. (i) Let $\mathbf{L}$ be the set of least-squares solutions of (4.4). Then $\mathbf{L}$ can written as follows

$$\mathbf{L} = \{u \in \mathcal{X} \,|\, \|Au - f\|_{\mathcal{Y}} \leqslant \mu\},$$

where $\mu := \inf\{\|Av - f\|_{\mathcal{Y}} | v \in \mathcal{X}\}$. Since A is $\tau_{\mathcal{X}} \to \tau_{\mathcal{Y}}$ continuous and $\|\cdot\|_{\mathcal{Y}}$ is $\tau_{\mathcal{Y}}$-l.s.c., $\mathbf{L}$ is $\tau_{\mathcal{X}}$-closed.

Consider the following problem

$$\inf_{u \in \mathbf{L}} \mathcal{J}(u) = \inf_{u \in \mathcal{X}} \mathcal{J}(u) + \chi_{\mathbf{L}}(u). \tag{4.7}$$

By the assumption that we made in the beginning of this section, this problem is feasible, i.e. there exists $u \in \mathbf{L}$ with $\mathcal{J}(u) < \infty$. The objective function in (4.7) is bounded from below. Using similar arguments as in Lemma 4.46, we conclude that it is also coercive. Since $\mathbf{L}$ is $\tau_{\mathcal{X}}$-closed, $\chi_{\mathbf{L}}$ is $\tau_{\mathcal{X}}$-l.s.c. By assumption ii, $\mathcal{J}$ is also $\tau_{\mathcal{X}}$-l.s.c. So, (4.7) satisfies the assumptions of the direct method (Theorem 4.15) and hence a minimiser exists.

(ii) From Lemma 4.46 we know that the objective function Φ_α in (4.6) is coercive. It is also bounded from below. Since $\mathcal{J}$ is $\tau_{\mathcal{X}}$-l.s.c., A is $\tau_{\mathcal{X}} \to \tau_{\mathcal{Y}}$ continuous and $\|\cdot\|_{\mathcal{Y}}$ is $\tau_{\mathcal{Y}}$-l.s.c., we get that Φ_α is $\tau_{\mathcal{X}}$-l.s.c. Using the direct method, we conclude that (4.6) has a minimiser.

□

Now we study the behaviour of the minimiser of (4.6) with $f = f_\delta$ (perturbed measurement) as $\delta \to 0$ when $\alpha = \alpha(\delta)$ is chosen according to an appropriate a priori parameter choice rule. For simplicity, we will do this in the case when $\inf\{\|Av - f\|_{\mathcal{Y}} | v \in \mathcal{X}\} = 0$, i.e. least-squares solutions are actually solutions of (4.4).

Theorem 4.48. *Let the assumptions of Theorem 4.47 hold and suppose that* $\inf\{\|Av - f\|_{\mathcal{Y}} | v \in \mathcal{X}\} = 0$. *Let* $\alpha = \alpha(\delta)$ *be such that*

$$\lim_{\delta\to 0} \alpha(\delta) = 0 \quad \textit{and} \quad \limsup_{\delta\to 0} \frac{\delta^2}{\alpha(\delta)} = 0$$

and suppose that for each $\delta > 0$, *there exists* $f_\delta \in \mathcal{Y}$ *such that* $\|f - f_\delta\| \leqslant \delta$. *Then* $u_\delta := u_\delta^{\alpha(\delta)} \overset{\tau_{\mathcal{X}}}{\to} u_{\mathcal{J}}^\dagger$ *as* $\delta \to 0$ *(possibly, along a subsequence) and* $\mathcal{J}(u_\delta) \to \mathcal{J}(u_{\mathcal{J}}^\dagger)$, *where* $u_{\mathcal{J}}^\dagger$ *is a* $\mathcal{J}$*-minimising solution.*

Proof. Let u_0 be any $\mathcal{J}$-minimising solution (which exists by Theorem 4.47). Since u_δ solves (4.6) with $\alpha = \alpha(\delta)$, we get that

$$\begin{aligned} \frac{1}{2}\|Au_\delta - f_\delta\|_{\mathcal{Y}}^2 + \alpha(\delta)\mathcal{J}(u_\delta) &\leqslant \frac{1}{2}\|Au_0 - f_\delta\|_{\mathcal{Y}}^2 + \alpha(\delta)\mathcal{J}(u_0) \\ &\leqslant \frac{\delta^2}{2} + \alpha(\delta)\mathcal{J}(u_0). \end{aligned} \tag{4.8}$$

Therefore, we have the following two estimates

$$\mathcal{J}(u_\delta) \leqslant \frac{\delta^2}{2\alpha(\delta)} + \mathcal{J}(u_0) \leqslant C, \tag{4.9a}$$

$$\|Au_\delta - f_\delta\|_{\mathcal{Y}} \leqslant \sqrt{\delta^2 + 2\alpha(\delta)\mathcal{J}(u_0)} \leqslant C', \tag{4.9b}$$

The right-hand side in (4.9a) is bounded uniformly in δ, because $\limsup_{\delta\to 0} \delta^2/\alpha(\delta) = 0$ by assumption and $\mathcal{J}(u_0)$ is a constant independent of δ. The right-hand side in (4.9b) is bounded, because $\mathcal{J}(u_0)$ is a constant and $\delta, \alpha(\delta) \to 0$.

Therefore, both $\mathcal{J}(u_\delta)$ and $\|Au_\delta - f_\delta\|_{\mathcal{Y}}$ are uniformly bounded. Proceeding in a similar way as in Lemma 4.46, we get that

$$\|u_\delta\| \leqslant C$$

for all δ. Now let $\delta_n \downarrow 0$ be an arbitrary null sequence. Since u_{δ_n} is bounded, it contains a $\tau_{\mathcal{X}}$-convergent subsequence (which we don't relabel)

$$u_{\delta_n} \overset{\tau_{\mathcal{X}}}{\to} u_{\mathcal{J}}^\dagger \quad \text{as } n \to \infty.$$

We will show that $u_{\mathcal{J}}^\dagger$ is a $\mathcal{J}$-minimising solution. From (4.9b) we observe that

$$\liminf_{n\to\infty} \|Au_{\delta_n} - f_{\delta_n}\|_{\mathcal{Y}} \leqslant \liminf_{n\to\infty} \sqrt{\delta_n^2 + 2\alpha(\delta_n)\mathcal{J}(u_0)} = 0.$$

Since A is $\tau_{\mathcal{X}} \to \tau_{\mathcal{Y}}$ continuous and $\|\cdot\|_{\mathcal{Y}}$ is $\tau_{\mathcal{Y}}$-l.s.c., we get that

$$\|Au_{\mathcal{J}}^\dagger - f\|_{\mathcal{Y}} \leqslant \liminf_{n\to\infty} \|Au_{\delta_n} - f\|_{\mathcal{Y}} \leqslant \liminf_{n\to\infty} \left(\|Au_{\delta_n} - f_{\delta_n}\|_{\mathcal{Y}} + \|f - f_{\delta_n}\|_{\mathcal{Y}}\right) = 0,$$

which shows that $u_{\mathcal{J}}^\dagger$ is a least-squares solution. Using the estimate (4.9a) and $\tau_{\mathcal{X}}$-lower semicontinuity of $\mathcal{J}$, we obtain

$$\mathcal{J}(u_{\mathcal{J}}^\dagger) \leqslant \liminf_{n\to\infty} \mathcal{J}(u_{\delta_n}) \leqslant \limsup_{n\to\infty} \mathcal{J}(u_{\delta_n}) \leqslant \limsup_{n\to\infty} \frac{\delta^2}{2\alpha(\delta)} + \mathcal{J}(u_0) = \mathcal{J}(u_0). \tag{4.10}$$

Since u_0 was an arbitrary $\mathcal{J}$-minimising solution and $\mathcal{J}(u_{\mathcal{J}}^\dagger) \leqslant \mathcal{J}(u_0)$, we conclude that $\mathcal{J}(u_{\mathcal{J}}^\dagger)$ is also a $\mathcal{J}$-minimising solution. Finally, since $\mathcal{J}(u_{\mathcal{J}}^\dagger) = \mathcal{J}(u_0)$, we conclude from (4.10) that

$$\liminf_{n\to\infty} \mathcal{J}(u_{\delta_n}) = \limsup_{n\to\infty} \mathcal{J}(u_{\delta_n}) = \lim_{n\to\infty} \mathcal{J}(u_{\delta_n}) = \mathcal{J}(u_{\mathcal{J}}^\dagger),$$

which completes the proof. □

Remark 4.49. The theorem proves convergence of the regularised solutions in $\tau_{\mathcal{X}}$, which may differ from the strong topology. However, if $\mathcal{J}$ satisfies the *Radon-Riesz property* with respect to the topology $\tau_{\mathcal{X}}$, i.e. $u_j \overset{\tau_{\mathcal{X}}}{\to} u$ and $\mathcal{J}(u_j) \to \mathcal{J}(u)$ imply $\|u_j - u\| \to 0$, then we get convergence in the norm topology. An example of a functional satisfying the Radon-Riesz property is the norm in a Hilbert (or reflexive Banach) space with $\tau_{\mathcal{X}}$ being the weak topology.

Examples of regularisers

Example 4.50. Let $\mathcal{X}$ be a Hilbert space and $\mathcal{J}(u) = \|u\|^2$. The norm in a Hilbert space is weakly l.s.c. By Theorem 4.2 we know that (norm) bounded sequences have weakly convergent subsequences. Therefore, Assumption (ii) of Theorem 4.47 is satisfied with $\tau_{\mathcal{X}}$ being the weak topology and we obtain weak convergence of the regularised solutions. However, since the norm in a Hilbert space has the Radon-Riesz property, we also get strong convergence. The same approach works in reflexive Banach spaces.

A classical example is regularisation in Sobolev spaces such as the space H^1 of L^2 functions whose weak derivatives are also in L^2. In the one-dimensional case, the space H^1 consists only of continuous functions (in higher dimensions it is true for Sobolev spaces with some other exponents), therefore, the regularised solutions will also be continuous. For this reason, the regulariser $\mathcal{J}(u) = \|u\|_{H^1}$ is sometimes referred to as the *smoothing functional*. Whilst desirable in some applications, in imaging smooth reconstructions are usually not favourable, since images naturally contain edges and therefore are not continuous functions. To overcome this issue, other regularisers have been introduced that we will discuss later.

Example 4.51 (ℓ^1-regularisation)**.** Let $\mathcal{X} = \ell^2$ be space of all square summable sequences (i.e. such that $\|u\|_{\ell^2}^2 = \sum_{i=1}^{\infty} u_i^2 < +\infty$). For example, u can represent the coefficients of a function in a basis (e.g., a Fourier basis or a wavelet basis). As a regularisation functional, let us use not the ℓ^2-norm, but the ℓ^1-norm:

$$\mathcal{J}(u) = \|u\|_{\ell^1} = \sum_{i=1}^{\infty} |u_i|.$$

By Example 4.5 $\mathcal{J}$ is weakly l.s.c. in ℓ^2. It is evident that $\ell^q \subset \ell^p$ and $\|\cdot\|_{\ell^p} \leqslant \|\cdot\|_{\ell^q}$ for $q \leqslant p$. Therefore, $\mathcal{J}(u) \leqslant C$ implies that $\|\cdot\|_{\ell^2} \leqslant C$ and, since ℓ^2 is a Hilbert space and bounded sequences have weakly convergent subsequences, we conclude that the sublevel sets of $\mathcal{J}(\cdot)$ are weakly sequentially compact in ℓ^2. Therefore, Assumption (ii) of Theorem 4.47 is satisfied with $\tau_{\mathcal{X}}$ being the weak topology in ℓ^2. Hence, we get weak convergence of regularised solutions in ℓ^2.

The motivation for using the ℓ^1-norm as the regulariser instead of the ℓ^2-norm is as follows. If the forward operator is non-injective, the inverse problem has more than one solution and the solutions form an affine subspace. In the context of sequence spaces representing coefficients of the solution in a basis, it is sometimes beneficial to look for solutions that are *sparse* in the sense that they have finite support, i.e. $|\operatorname{supp}(u)| < \infty$ with $\operatorname{supp}(u) = \{i \in \mathbf{N} \mid u_i \neq 0\}$. This allows explaining the signal with a finite (and often relatively small) number of basis functions and has widely ranging applications in, for instance, compressed sensing. A finite dimensional illustration of the sparsity of ℓ^1-regularised solutions is given in Figure 4.8. The corresponding minimisation problem

$$\min_{u \in \ell^2} \left\{ \frac{1}{2} \|Au - f\|_{\ell^2}^2 + \alpha \|u\|_1 \right\}. \tag{4.11}$$

is also called *LASSO* in the statistical literature.

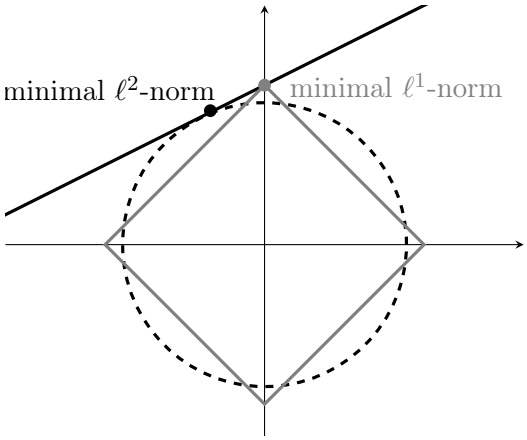

Figure 4.8: Non-injective operators have a non-trivial kernel, so that the inverse problem has more than one solution and the solutions form an affine subspace visualised by the solid line. Different regularisation functionals favour different solutions. The circle and the diamond indicate all points with constant ℓ^2-norm, respectively ℓ^1-norm, and the minimal ℓ^2-norm and ℓ^1-norm solutions are the intersections of the line with the circle, respectively the diamond. As it can be seen, the minimal ℓ^2-norm solution has two non-zero components while the minimal ℓ^1-norm solution has only one non-zero component and thus is *sparser*.

Example 4.52 (Elastic net regularisation)**.** The ℓ^1-regulariser described in the previous example sometimes delivers undesirable results for problems where there are highly correlated features and we need to identify all relevant ones, e.g. microarray data analysis (analysis of genomic sequences), in that it tends to select only one feature out of the relevant group instead of all relevant features of the group, i.e. it fails to identify the group structure. Elastic net regularisation helps to overcome this issue. The elastic net regulariser $\mathcal{J} : \ell^2 \to \bar{\mathbf{R}}_+$ is defined as follows

$$\mathcal{J}(u) := \alpha\|u\|_{\ell^1} + \beta\|u\|_{\ell^2}^2,$$

where $\alpha, \beta > 0$ are constants that balance the influence of the two terms. Since $\mathcal{J}$ is the sum of a 1-homogeneous term and a 2-homogeneous term, it satisfies assumptions of Lemma 4.45.

4.3 Total Variation Regularisation

As pointed out in Example 4.50, in imaging we are interested in regularisers that allow for discontinuities while maintaining sufficient regularity of the reconstructions. One popular choice is the so-called *total variation* regulariser [10].

Definition 4.53. *Let $\Omega \subset \mathbf{R}^n$ be a bounded domain and $u \in L^1(\Omega)$. Let $\mathcal{D}(\Omega, \mathbf{R}^n)$ be the following set of vector-valued* test functions *(i.e. functions that map from Ω to $\mathbf{R}^n$)*

$$\mathcal{D}(\Omega, \mathbf{R}^n) := \{\varphi \in C_0^\infty(\Omega; \mathbf{R}^n) | \sup_{x\in\Omega} \|\varphi(x)\|_2 \leqslant 1\}.$$

Total variation of $u \in L^1(\Omega)$ is defined as follows

$$\mathrm{TV}(u) = \sup_{\varphi\in\mathcal{D}(\Omega,\mathbf{R}^n)} \int_\Omega u(x) \operatorname{div} \varphi(x)\, \mathrm{d}x.$$

Remark 4.54. Definition 4.53 may seem a bit strange at the first glance, but we note that for a function $u \in L^1(\Omega)$ whose weak derivative ∇u exists and is also in $L^1(\Omega, \mathbf{R}^n)$ (i.e. u belongs to the Sobolev space $W^{1,1}(\Omega)$) we obtain, integrating by parts, that

$$\mathrm{TV}(u) = \sup_{\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)} \int_\Omega -\langle \nabla u(x), \varphi(x) \rangle \, \mathrm{d}x.$$

By the Cauchy–Schwarz inequality we get that $|\langle \nabla u(x), \varphi(x) \rangle| \leqslant \|\nabla u(x)\|_2 \|\varphi(x)\|_2 \leqslant \|\nabla u(x)\|_2$ for a.e. $x \in \Omega$. On the other hand, choosing φ such that $\varphi(x) = -\frac{\nabla u(x)}{\|\nabla u(x)\|_2}$ (technically, such φ is not necessarily in $\mathcal{D}(\Omega, \mathbf{R}^n)$, but we can approximate it with functions from $\mathcal{D}(\Omega, \mathbf{R}^n)$, since any function in $W^{1,1}(\Omega)$ can be approximated with smooth functions [2, Theorem 3.17]; we omit the technicalities here), we get that $-\langle \nabla u(x), \varphi(x) \rangle = \|\nabla u(x)\|_2$. Therefore, the supremum over $\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)$ is equal to

$$\mathrm{TV}(u) = \int_\Omega \|\nabla u(x)\|_2 \, \mathrm{d}x = \|\nabla u\|_{L^1}.$$

This shows that TV just penalises the the L^1 norm (of the pointwise 2-norm) of the gradient for any $u \in W^{1,1}(\Omega)$. However, we will see that the space of functions that have finite value of TV is larger than $W^{1,1}(\Omega)$ and contains, for instance, discontinuous functions.

Remark 4.55. It can be shown [8] that for any $u \in L^1(\Omega)$

$$\mathrm{TV}(u) = \|\nabla u\|_{\mathfrak{M}},$$

where ∇ is the distributional gradient and $\|\cdot\|_{\mathfrak{M}}$ is the Radon norm. That is, Total Variation extends the L^1 norm of the gradient for functions whose gradient is not a Lebesgue-measurable function. We will not use this interpretation of the Total Variation to simplify the presentation and refer the interested reader to [8] for details.

Proposition 4.56. TV *is a proper, convex and absolutely* 1*-homogeneous functional* $L^1(\Omega) \to \bar{\mathbf{R}}$. *For any constant function* $\mathbf{c}\colon \mathbf{c}(x) \equiv c \in \mathbf{R}$ *for all* x *and any* $u \in L^1(\Omega)$

$$\mathrm{TV}(\mathbf{c}) = 0 \quad \textit{and} \quad \mathrm{TV}(u + \mathbf{c}) = \mathrm{TV}(u).$$

Proof. Left as exercise. □

Remark 4.57. It can be shown that the opposite implication holds, i.e. $\mathrm{TV}(u) = 0$ implies that u is constant. in other words,

$$\ker(\mathrm{TV}) = \{u \in L^1(\Omega) | u(x) \equiv c\}. \tag{4.12}$$

The easiest way to see this is using the Radon measure interpretation in Remark 4.55. Because of time constraints, we will omit the proof.

Example 4.58 (TV of an indicator function)**.** Suppose $C \subset \Omega \subset \mathbf{R}^2$ is a bounded domain with smooth boundary and $u(\cdot) = \mathbf{1}_C(\cdot)$ is its indicator function, i.e.

$$\mathbf{1}_C(x) = \begin{cases} 1 & x \in C, \\ 0 & x \in X \setminus C. \end{cases}$$

Then, using the divergence theorem, we get that for any test function $\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)$

$$\int_\Omega u(x) \operatorname{div} \varphi(x) \, \mathrm{d}x = \int_C \operatorname{div} \varphi(x) \, \mathrm{d}x = \int_{\partial C} \langle \varphi(x), \mathbf{n}_{\partial C}(x) \rangle \, \mathrm{d}l,$$

where ∂C is the boundary of C and $\mathbf{n}_{\partial C}(x)$ is the unit normal at x. Hence,

$$\begin{aligned}\mathrm{TV}(u) &= \sup_{\varphi\in\mathcal{D}(\Omega,\mathbf{R}^n)} \int_\Omega u(x)\,\mathrm{div}\,\varphi(x)\,\mathrm{d}x = \sup_{\varphi\in\mathcal{D}(\Omega,\mathbf{R}^n)} \int_{\partial C} \langle \varphi(x), \mathbf{n}_{\partial C}(x)\rangle\,\mathrm{d}l \\ &\leqslant \sup_{\varphi\in\mathcal{D}(\Omega,\mathbf{R}^n)} \int_{\partial C} \|\varphi(x)\|\|\mathbf{n}_{\partial C}(x)\|\,\mathrm{d}l \leqslant \sup_{\varphi\in\mathcal{D}(\Omega,\mathbf{R}^n)} \int_{\partial C} \mathrm{d}l = \mathrm{Per}(C),\end{aligned}$$

where $\mathrm{Per}(C)$ is the perimeter of C. On the other hand, since ∂C is smooth and $\|\mathbf{n}_{\partial C}(x)\| = 1$ for every x, $\mathbf{n}_{\partial C}$ can be extended to feasible vector field on Ω (i.e. one that is in $D(\Omega, \mathbf{R}^n)$). Therefore, with φ one such vector field, we get that

$$\mathrm{TV}(u) \geqslant \int_{\partial C} \langle \varphi(x), \mathbf{n}_{\partial C}(x)\rangle\,\mathrm{d}l \geqslant \int_{\partial C} \|\mathbf{n}_{\partial C}(x)\|^2\,\mathrm{d}l = \int_{\partial C} 1\cdot\,\mathrm{d}l = \mathrm{Per}(C),$$

Therefore, $\mathrm{TV}(\mathbf{1}_C) = \mathrm{Per}(C)$ for any domain with smooth boundary. This can be extended to domains with Lipschitz boundary by constructing a sequence of functions in $D(\Omega, \mathbf{R}^n)$ that converge pointwise to $\mathbf{n}_{\partial C}$.

We now study properties of functions that have a finite value of TV.

Definition 4.59. *The functions $u \in L^1(\Omega)$ with a finite value of* TV *form a normed space called the space of functions of bounded variation (the* BV*-space) defined as follows*

$$\mathrm{BV}(\Omega) := \{u \in L^1(\Omega) \mid \|u\|_{\mathrm{BV}} := \|u\|_{L^1} + \mathrm{TV}(u) < \infty\}.$$

Remark 4.60. It can be shown that the space BV is the dual of a separable Banach space [8] and that weak-* convergence $u_n \rightharpoonup^* u$ in BV is equivalent to strong convergence $u_n \to u$ in L^1 and convergence of the values $\mathrm{TV}(u_n) \to \mathrm{TV}(u)$. The proof is outside the scope of these notes.

We note that $\mathrm{BV}(\Omega)$ is compactly embedded in $L^1(\Omega)$. We start with the following classical result.

Theorem 4.61 (Rellich-Kondrachov, [2, Theorem 6.3])**.** *Let $\Omega \subset \mathbf{R}^n$ be a bounded Lipschitz domain (i.e. non-empty, open, connected and with Lipschitz boundary) and $p, m \in \mathbf{N}$. Let*

$$p^* := \begin{cases} \frac{np}{n-mp} & \text{if } n > mp, \\ \infty & \text{if } n \leqslant mp. \end{cases}$$

Then the embedding $W^{m,p}(\Omega) \to L^q(\Omega)$ is continuous for all $1 \leqslant q \leqslant p^*$ and compact for all $1 \leqslant q < p^*$.

Since functions from $\mathrm{BV}(\Omega)$ can be approximated by functions in the Sobolev space $W^{1,1}(\Omega)$ [4, Theorem 3.9], the Rellich-Kondrachov Theorem (with $p = 1$, $m = 1$) gives us the following

Corollary 4.62 ([4, Corollary 3.49])**.** For any bounded Lipschitz domain $\Omega \subset \mathbf{R}^n$, the embedding

$$\mathrm{BV}(\Omega) \Subset L^1(\Omega)$$

is compact for any $n \geqslant 2$ and the embedding

$$\mathrm{BV}(\Omega) \hookrightarrow L^2(\Omega)$$

is continuous for $n = 2$.

Now we will show that TV is lower-semicontinuous in L^1.

Theorem 4.63. *Let $\Omega \subset \mathbf{R}^n$ be open and bounded. Then the total variation is l.s.c. in $L^1(\Omega)$.*

Proof. Let $\{u_j\}_{j\in\mathbf{N}} \subset \mathrm{BV}(\Omega)$ be a sequence converging in $L^1(\Omega)$ with $u_j \to u$ in $L^1(\Omega)$. Then for any test function $\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)$ we have that

$$\int_\Omega u_j(x) \operatorname{div}\varphi(x)\,\mathrm{d}x \to \int_\Omega u(x)\operatorname{div}\varphi(x)\,\mathrm{d}x.$$

Indeed, note that $\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)$ implies that $\|\operatorname{div}\varphi\|_\infty < \infty$, so that

$$\int_\Omega |(u_j(x) - u(x))\operatorname{div}\varphi(x)|\,\mathrm{d}x \leqslant \|u_j - u\|_1 \|\mathrm{div}\varphi\|_\infty \to 0$$

by Hölder's inequality. Now, for any $j \in \mathbf{N}$, we have

$$\int_\Omega u_j(x)\operatorname{div}\varphi(x)\,\mathrm{d}x \leqslant \sup_{\psi\in\mathcal{D}(\Omega,\mathbf{R}^n)} \int_\Omega u_j(x)\operatorname{div}\psi(x)\,\mathrm{d}x = \mathrm{TV}(u_j),$$

so that

$$\begin{aligned}\int_\Omega u(x)\operatorname{div}\varphi(x)\,\mathrm{d}x &= \lim_{j\to\infty}\int_\Omega u_j(x)\operatorname{div}\varphi(x)\,\mathrm{d}x \\ &\leqslant \liminf_{j\to\infty}\mathrm{TV}(u_j).\end{aligned}$$

Taking the supremum over $\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)$, we find that $\mathrm{TV}(u) \leqslant \liminf_{j\to\infty}\mathrm{TV}(u_j)$, i.e. TV is lower semi-continuous with respect to the L^1-norm. □

Since the null space of total variation (4.12) is non-trivial, TV cannot be coercive on L^1. However, the following result helps.

Proposition 4.64 ([4, Remark 3.50])**.** *Let $\Omega \subset \mathbf{R}^n$ be a bounded Lipschitz domain. Then there exists a constant $C > 0$ such that for all $u \in \mathrm{BV}(\Omega)$ the* Poincaré *inequality is satisfied*

$$\|u - u_\Omega\|_{L^1} \leqslant C\,\mathrm{TV}(u),$$

where $u_\Omega := \frac{1}{|\Omega|}\int_\Omega u(x)\,\mathrm{d}x$ is the mean value of u over Ω.

Corollary 4.65. It is often useful to consider a subspace $\mathrm{BV}_0(\Omega) \subset \mathrm{BV}(\Omega)$ of functions with zero mean, i.e.

$$\mathrm{BV}_0(\Omega) := \left\{u \in \mathrm{BV}(\Omega)\middle|\int_\Omega u(x)\,\mathrm{d}x = 0\right\}. \tag{4.13}$$

Then for every function $u \in \mathrm{BV}_0(\Omega)$ we have that

$$\|u\|_{L^1} \leqslant C\,\mathrm{TV}(u).$$

Clearly, $\mathrm{BV}_0 \subset L^1_0 := \{u \in L^1 | \int_\Omega u(x)\,\mathrm{d}x = 0\}$ and TV is coercive on this subspace. Since $\dim(\ker(\mathrm{TV})) = 1 < \infty$, we have

$$L^1 = L^1_0 \oplus \ker(\mathrm{TV}).$$

Combining all the above results we get

Theorem 4.66. *Let $\mathcal{X} = L^1(\Omega)$, where $\Omega \subset \mathbf{R}^n$ is bounded Lipschitz, and $\mathcal{Y}$ be a Banach space. Let $A\colon L^1 \to \mathcal{Y}$ be a linear bounded operator such that $A\mathbf{1} \neq 0$, where $\mathbf{1}$ is the constant-one function. Then minimisers of the following problem*

$$\min_{u \in L^1(\Omega)} \frac{1}{2}\|Au - f_\delta\|_{\mathcal{Y}}^2 + \alpha(\delta)\,\mathrm{TV}(u)$$

converge strongly in L^1 to a TV*-minimising solution as $\delta \to 0$ if $\alpha(\delta)$ is chosen as required by Theorem 4.48.*

Proof. We have established all ingredients required for Theorem 4.48 to hold except that bounded sequences in L^1 may not have convergent subsequences (L^1 is not a dual space). However, the compact embedding from Corollary 4.62 guarantees that sequences with a bounded value of TV have subsequences that converge strongly in L^1. □

Remark 4.67. One can replace optimisation over $u \in L^1$ with optimisation over $u \in \mathrm{BV}$, which is the effective domain of the objective function.

Total Variation is widely used in imaging applications [20]. The so-called Rudin–Osher–Fatemi (ROF) model for image denoising [16] consists in minimising the following functional

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2}\|Iu - f_\delta\|_{L^2(\Omega)}^2 + \alpha\,\mathrm{TV}(u), \tag{4.14}$$

where $\Omega \subset \mathbf{R}^2$. In this case, the forward operator I is the embedding operator $\mathrm{BV}(\Omega) \to L^2(\Omega)$, which is continuous for two-dimensional domains (see Corollary 4.62). Clearly, $A\mathbf{1} \neq 0$ is satisfied. More generally, one considers the following optimisation problem

$$\min_{u \in \mathrm{BV}(\Omega)} \|Au - f_\delta\|_2^2 + \alpha\,\mathrm{TV}(u), \tag{4.15}$$

where $A\colon \mathrm{BV}(\Omega) \to L^2(\Omega)$ is such that $A\mathbf{1} \neq 0$.

Chapter 5

Convex Duality

In Chapter 4 we have established convergence of a regularised solution u_δ to a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ as $\delta \to 0$. However, we didn't get any results on the *speed* of this convergence, which is referred to as the *convergence rate*.

In modern regularisation methods, convergence rates are usually studied using *Bregman divergences* associated with the (convex) regularisation functional $\mathcal{J}$. Recall that for a convex functional $\mathcal{J}$, $u, v \in \mathcal{X}$ such that $\mathcal{J}(v) < \infty$ and $q \in \partial\mathcal{J}(v)$, the (generalised) Bregman divergence is given by the following expression (cf. Def. 4.33)

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

Also widely used is the *symmetric* Bregman divergence (cf. Def. 4.35) given by the following expression (here $p \in \partial\mathcal{J}(u)$)

$$D^{\text{symm}}_{\mathcal{J}}(u, v) = D^q_{\mathcal{J}}(u, v) + D^p_{\mathcal{J}}(v, u) = \langle p - q, u - v\rangle .$$

Bregman divergences appear to be a natural distance measure between a regularised solution u_δ and a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$. For instance, for classical Hilbert space regularisation with $\mathcal{J}(u) = \frac{1}{2}\|u\|^2_{\mathcal{X}}$, the subgradient at $u^\dagger_{\mathcal{J}}$ is $p_{u^\dagger_{\mathcal{J}}} = u^\dagger_{\mathcal{J}}$ (since $\mathcal{J}$ is differentiable) and we get the following expression

$$\begin{aligned} D^{u^\dagger_{\mathcal{J}}}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) &= \frac{1}{2}\|u_\delta\|^2_{\mathcal{X}} - \frac{1}{2}\|u^\dagger_{\mathcal{J}}\|^2_{\mathcal{X}} - \langle u^\dagger_{\mathcal{J}}, u_\delta - u^\dagger_{\mathcal{J}}\rangle \\ &= \frac{1}{2}(\|u_\delta\|^2_{\mathcal{X}} - 2\langle u^\dagger_{\mathcal{J}}, u_\delta\rangle + \|u^\dagger_{\mathcal{J}}\|^2_{\mathcal{X}}) = \frac{1}{2}\|u_\delta - u^\dagger_{\mathcal{J}}\|^2_{\mathcal{X}}, \end{aligned}$$

which happens to coincide with the symmetric Bregman divergence. Therefore, in the classical L^2-case, the Bregman divergence just measures the L^2-distance between a regularised solution and a $\mathcal{J}$-minimising solution. As we have seen in an examples sheet, subgradients of absolutely one-homogeneous functional carry structural information about the solution such as locations of non-zero components of a vector $u^\dagger_{\mathcal{J}} \in \ell^1$.

We are looking for a convergence rate of the following form

$$D^{\text{symm}}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) \leqslant \psi(\delta),$$

where $\psi\colon \mathbf{R}_+ \to \mathbf{R}_+$ is a known function of δ such that $\psi(\delta) \to 0$ as $\delta \to 0$.

5.1 Duality in convex optimisation

Consider the following optimisation problem

$$\inf_{u\in\mathcal{X}} E(Au) + F(u), \tag{$\mathcal{P}$}$$

where $E\colon \mathcal{Y} \to \bar{\mathbf{R}}$ and $F\colon \mathcal{X} \to \bar{\mathbf{R}}$ are proper, convex and lower semicontinuous functions and $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ is a linear bounded operator. Since E is convex and lower semicontinuous, it can be written as the convex conjugate of its conjugate E^*:

$$E(y) = \sup_{\eta\in\mathcal{Y}^*} \langle \eta, y\rangle - E^*(\eta) \quad y \in \mathcal{Y}.$$

Hence, we can rewrite $(\mathcal{P})$ as follows

$$\inf_{u\in\mathcal{X}} \sup_{\eta\in\mathcal{Y}^*} \langle \eta, Au\rangle - E^*(\eta) + F(u). \tag{$\mathcal{S}$}$$

This problem is referred to as the *saddle point problem*, whereas $(\mathcal{P})$ is referred to as the *primal problem*. Since inf sup $\geqslant$ sup inf always holds, we get that

$$\begin{aligned}
\inf_{u\in\mathcal{X}} E(Au) + F(u) &\geqslant \sup_{\eta\in\mathcal{Y}^*} \inf_{u\in\mathcal{X}} \langle \eta, Au\rangle - E^*(\eta) + F(u) \\
&= \sup_{\eta\in\mathcal{Y}^*} \inf_{u\in\mathcal{X}} \langle A^*\eta, u\rangle - E^*(\eta) + F(u) \\
&= \sup_{\eta\in\mathcal{Y}^*} \left\{ -E^*(\eta) - \sup_{u\in\mathcal{X}} \left[\langle -A^*\eta, u\rangle - F(u)\right] \right\} \\
&= \sup_{\eta\in\mathcal{Y}^*} -E^*(\eta) - F^*(-A^*\eta).
\end{aligned}$$

The last problem

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

is called the *dual problem*. The fact that the optimal value of the primal is always less than or equal to the optimal value of the dual problem is referred to as *weak duality* and the difference between these two optimal values is referred to as the *duality gap*. Whenever the two optimal values are in fact equal, one speaks of *strong duality*. Sufficient conditions for strong duality are given by the following theorem.

Theorem 5.1 ([11, Chapter III Theorem 4.1 and Remark 4.2])**.** *Suppose that*

(i) the function $E(Au) + F(u)\colon \mathcal{X} \to \bar{\mathbf{R}}$ is proper, convex, l.s.c. and coercive;

(ii) $\exists u_0 \in \mathcal{X}$ s.t. $F(u_0) < +\infty$, $E(Au_0) < +\infty$ and $E(y)$ is continuous at $y = Au_0$.

Then

(i) The dual problem $(\mathcal{D})$ has at least one solution $\widehat{\eta}$;

(ii) There is no duality gap between $(\mathcal{P})$ and $(\mathcal{D})$, i.e. strong duality holds;

(iii) If $(\mathcal{P})$ has an optimal solution $\widehat{u}$, then the following optimality conditions hold

$$-A^*\widehat{\eta} \in \partial F(\widehat{u}), \quad \widehat{\eta} \in \partial E(A\widehat{u}).$$

Note that existence of a primal solution is *not* guaranteed by this theorem.

5.2 The dual problem of the variational regularisation problem

Recall that u_δ solves the following problem

$$\min_{u\in\mathcal{X}} \frac{1}{2}\|Au - f_\delta\|_{\mathcal{Y}}^2 + \alpha\mathcal{J}(u). \tag{5.1}$$

with an appropriately chosen $\alpha = \alpha(\delta)$, where $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces, $A \in \mathcal{L}(\mathcal{X},\mathcal{Y})$ and $E\colon \mathcal{Y} \to \bar{\mathbf{R}}$ and $\mathcal{J}\colon \mathcal{X} \to \bar{\mathbf{R}}$ is proper, convex and l.s.c. and satisfies the assumptions of Theorem 4.47. For simplicity of presentation, we will also assume throughout this chapter that $\mathcal{J}$ is absolutely one-homogeneous and that $\inf\{\|Av-f\|\colon v\in\mathcal{X}\} = 0$, i.e. $Au_{\mathcal{J}}^\dagger = f$ for any $\mathcal{J}$-minimising solution.

To apply the results of the previous section to (5.1), we take (in the notation of the previous section)

$$E(y) := \frac{1}{2}\|y - f_\delta\|_{\mathcal{Y}}^2, \quad F(u) := \alpha\mathcal{J}(u).$$

Lemma 5.2. *Let X be a Banach space with norm $\|\cdot\|_X$ and let $\|\cdot\|_{X^*}$ be the norm on the dual space of X. Let $\varphi(x) := \frac{1}{2}\|x\|_X^2$. Then the convex conjugate of φ is*

$$\varphi^*(\xi) = \frac{1}{2}\|\xi\|_{X^*}^2, \quad \xi \in X^*.$$

Proof. This is an exercise on one of the examples sheets. □

Corollary 5.3. Theorem 4.32 implies that for any $x \in X$ and any $\xi \in \partial\varphi(x)$ it holds

$$\frac{1}{2}\|x\|_X^2 + \frac{1}{2}\|\xi\|_{X^*}^2 = \langle \xi, x\rangle.$$

Using the "Cauchy–Schwarz inequality" (really just the definition of the norm on the dual space) on the right-hand side and rearranging terms, we get that $(\|x\|_X - \|\xi\|_{X^*})^2 = 0$ and hence

$$\|\xi\|_{X^*} = \|x\|_X.$$

Now, for E and F as defined above, we get

$$\begin{aligned} E^*(\eta) &= \sup_{f\in\mathcal{Y}} \langle \eta, f\rangle - \frac{1}{2}\|f - f_\delta\|_{\mathcal{Y}}^2 = \langle \eta, f_\delta\rangle - \sup_{g\in\mathcal{Y}} \left(\langle \eta, g\rangle - \frac{1}{2}\|g\|_{\mathcal{Y}}^2\right) = \langle \eta, f_\delta\rangle + \frac{1}{2}\|\eta\|_{\mathcal{Y}^*}^2, \\ F^*(p) &= \chi_{\partial\mathcal{J}(0)}\left(\frac{p}{\alpha}\right), \end{aligned}$$

where the second equality holds since F is absolutely one-homogeneous. Hence, the dual problem of (5.1) is given by

$$\sup_{\eta\in\mathcal{Y}^*} -\langle \eta, f_\delta\rangle - \frac{1}{2}\|\eta\|_{\mathcal{Y}^*}^2 - \chi_{\partial\mathcal{J}(0)}\left(\frac{-A^*\eta}{\alpha}\right).$$

Let us rewrite this in a slightly more convenient form. Denote $\mu := -\frac{\eta}{\alpha} \in \mathcal{Y}^*$. Since $-\chi_{\partial\mathcal{J}(0)} = -\infty$ outside $\partial\mathcal{J}(0)$, we get the following equivalent problem

$$\sup_{\substack{\mu\in\mathcal{Y}^* \\ A^*\mu\in\partial\mathcal{J}(0)}} \alpha\left(\langle \mu, f_\delta\rangle - \frac{\alpha}{2}\|\mu\|_{\mathcal{Y}}^2\right). \tag{5.2}$$

Let us check if the assumptions of Theorem 5.1 are satisfied. Condition (i) (coercivity) is guaranteed by Lemma 4.46. Condition (ii) (continuity of E) is satisfied at $u_0 = 0$. Therefore, for any $\delta > 0$ there exists a solution μ_δ of the dual problem (5.2).

Existence of a primal solution u_δ is guaranteed by Theorem 4.47. Indeed, let us take $\tau_\mathcal{X}$ to be the weak-* topology in $\mathcal{X}$ and $\tau_\mathcal{Y}$ a topology in $\mathcal{Y}$ such that A is $\tau_\mathcal{X}$-$\tau_\mathcal{Y}$ continuous and the norm in $\mathcal{Y}$ is $\tau_\mathcal{Y}$-l.s.c. (weak-*, weak or strong topologies will work). For example, if $\mathcal{Y}$ has a separable predual, we can take $\tau_\mathcal{Y}$ to be the weak-* topology on $\mathcal{Y}$. It can be easily verified that A is weak-*-weak-* continuous if it is the dual of another operator $A = B^*$ (where B acts from the predual of $\mathcal{Y}$ into the predual of $\mathcal{X}$). With these choices, the conditions of Theorem 4.47 are satisfied. Hence, by strong duality we have that

$$\frac{1}{2}\|Au_\delta - f_\delta\|_\mathcal{Y}^2 + \alpha\mathcal{J}(u_\delta) = \alpha\langle\mu_\delta, f_\delta\rangle - \frac{\alpha^2}{2}\|\mu_\delta\|_\mathcal{Y}^2.$$

Optimality conditions (iii) from Theorem 5.1 take the following form

$$A^*\mu_\delta \in \partial\mathcal{J}(u_\delta), \quad -\alpha\mu_\delta \in \partial\left(\frac{1}{2}\|\cdot\|_\mathcal{Y}^2\right)(Au_\delta - f_\delta). \tag{5.3}$$

From Corollary 5.3 we conclude that

$$\|\alpha\mu_\delta\|_{\mathcal{Y}^*} = \|Au_\delta - f_\delta\|_\mathcal{Y}. \tag{5.4}$$

Also, comparing the values of $\frac{1}{2}\|\cdot\|^2$ at 0 and at $Au_\delta - f_\delta$ and using the fact that $-\alpha\mu_\delta$ is a subgradient, we get that

$$0 \geqslant \frac{1}{2}\|Au_\delta - f_\delta\|_\mathcal{Y}^2 + \langle -\alpha\mu_\delta, 0 - (Au_\delta - f_\delta)\rangle$$

and therefore

$$\langle\alpha\mu_\delta, Au_\delta - f_\delta\rangle \leqslant -\frac{1}{2}\|Au_\delta - f_\delta\|_\mathcal{Y}^2. \tag{5.5}$$

We will use the estimates (5.4) and (5.5) later in Theorem 5.7.

5.3 Source Condition and Convergence Rates

Formal limits of problems (5.1) and (5.2) at $\delta = 0$ are

$$\inf_{u\colon Au=f} \mathcal{J}(u) = \inf_{u\in\mathcal{X}} \chi_{\{f\}}(Au) + \mathcal{J}(u) \tag{5.6}$$

and

$$\begin{aligned}\sup_{\mu\colon A^*\mu\in\partial\mathcal{J}(0)} \langle\mu, f\rangle &= \sup_{\mu\colon A^*\mu\in\partial\mathcal{J}(0)} \langle\mu, Au_\mathcal{J}^\dagger\rangle \\ &= \sup_{\mu\colon A^*\mu\in\partial\mathcal{J}(0)} \langle A^*\mu, u_\mathcal{J}^\dagger\rangle = \sup_{v\in\mathrm{im}(A^*)\cap\partial\mathcal{J}(0)} \langle v, u_\mathcal{J}^\dagger\rangle.\end{aligned} \tag{5.7}$$

Since the characteristic function $\chi_{\{f\}}(\cdot)$ is not continuous anywhere in its domain, Theorem 5.1 does not apply and we cannot guarantee that a solution of the dual limit problem (5.7) exists. Indeed, since $\mathrm{im}(A^*)$ is not closed, a solution may not exist. We will see that existence is guaranteed by the following condition, known as the source condition:

Definition 5.4 (Source condition [9]). *We say that a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ satisfies the* source condition *if*

$$\exists \mu^\dagger \in \mathcal{Y}^* \quad \text{such that} \quad A^*\mu^\dagger \in \partial \mathcal{J}(u^\dagger_{\mathcal{J}}), \tag{5.8}$$

i.e. if $\operatorname{im}(A^*) \cap \partial \mathcal{J}(u^\dagger_{\mathcal{J}}) \neq \emptyset$.

Remark 5.1. If $\mu^\dagger$ satisfies (5.8), then $\mu^\dagger$ is in fact a solution of the limiting dual problem (5.7): since $\mathcal{J}$ is absolutely 1-homogeneous, we have (by Proposition 4.40) $A^*\mu^\dagger \in \partial \mathcal{J}(0)$ and $\mathcal{J}(u^\dagger) = \langle A^*\mu^\dagger, u^\dagger \rangle = \langle \mu^\dagger, f \rangle$, where the right hand side is just the objective function of the limiting dual problem. Since $u^\dagger$ is a $\mathcal{J}$-minimising solution, $u^\dagger$ is feasible for the limiting primal problem (5.6), showing that in fact $\mu^\dagger$ is a solution of the dual problem.

First, we will see that the source condition is necessary for the dual solution μ_δ from (5.3) to stay bounded as $\delta \to 0$.

Theorem 5.5 (Necessary conditions). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $\mathcal{Y}$ separable. Let the conditions of Theorem 4.47 be satisfied and $\alpha = \alpha(\delta)$ be chosen as required by Theorem 4.48. Suppose that the dual solution μ_δ is bounded uniformly in δ. Then there exists $\mu^\dagger \in \mathcal{Y}^*$ such that $A^*\mu^\dagger \in \partial \mathcal{J}(u^\dagger_{\mathcal{J}})$.*

Proof. Consider an arbitrary sequence $\delta_n \downarrow 0$. Since $\|\mu_\delta\|_{\mathcal{Y}^*} \leqslant C$ for all δ, by the Banach–Alaoglu theorem we get that there exists a weakly-* convergent subsequence (that we do not relabel), i.e.

$$\mu_{\delta_n} \rightharpoonup^* \mu^\dagger \in \mathcal{Y}^*.$$

Then we get that

$$A^*\mu_{\delta_n} \rightharpoonup^* A^*\mu^\dagger.$$

Since $\partial \mathcal{J}(0)$ is weakly-* closed (Theorem 4.28) and $A^*\mu_{\delta_n} \in \partial \mathcal{J}(0)$ by (5.3), we get that

$$A^*\mu^\dagger \in \partial \mathcal{J}(0).$$

Since $\mathcal{J}$ is absolutely 1-homogeneous, we get by Proposition 4.37 that

$$\langle A^*\mu_{\delta_n}, u_{\delta_n} \rangle = \mathcal{J}(u_{\delta_n}) \to \mathcal{J}(u^\dagger_{\mathcal{J}}), \tag{5.9}$$

where convergence follows from Theorem 4.48. We also observe that

$$\begin{aligned} |\langle A^*\mu_\delta, u_\delta \rangle - \langle A^*\mu^\dagger, u^\dagger_{\mathcal{J}} \rangle| &= |\langle A^*\mu_\delta, u_\delta - u^\dagger_{\mathcal{J}} \rangle - \langle A^*(\mu^\dagger - \mu_\delta), u^\dagger_{\mathcal{J}} \rangle| \\ &\leqslant |\langle \mu_\delta, Au_\delta - f \rangle| + |\langle \mu^\dagger - \mu_\delta, f \rangle| \\ &\leqslant \|\mu_\delta\| \|Au_\delta - f\| + |\langle \mu^\dagger - \mu_\delta, f \rangle| \to 0, \end{aligned}$$

since $\|\mu_{\delta_n}\|_{\mathcal{Y}^*}$ is bounded, $\|Au_{\delta_n} - f\|_{\mathcal{Y}} \to 0$ and $\mu_{\delta_n} \rightharpoonup^* \mu^\dagger$. Combining this with (5.9), we get that

$$\mathcal{J}(u^\dagger_{\mathcal{J}}) = \langle A^*\mu^\dagger, u^\dagger_{\mathcal{J}} \rangle.$$

Since $A^*\mu^\dagger \in \partial \mathcal{J}(0)$ and $\mathcal{J}(u^\dagger_{\mathcal{J}}) = \langle A^*\mu^\dagger, u^\dagger_{\mathcal{J}} \rangle$, we conclude, using Proposition 4.40, that $A^*\mu^\dagger \in \partial \mathcal{J}(u^\dagger_{\mathcal{J}})$. □

So, the source condition is necessary for the boundedness of the dual solutions μ_δ as $\delta \to 0$. It turns out to also be sufficient.

Theorem 5.6 (Sufficient conditions)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $\mathcal{Y}$ separable. Let conditions of Theorem 4.47 be satisfied and $\alpha = \alpha(\delta)$ be chosen as required by Theorem 4.48. Suppose that the source condition* (5.8) *is satisfied at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$, and that the parameter choice rule is chosen so that α/δ is bounded. Then μ_δ is bounded uniformly in δ. Moreover, $\mu_\delta \rightharpoonup^* \mu^\dagger$ in $\mathcal{Y}^*$ as $\delta \to 0$ (perhaps, up to a subsequence), where $\mu^\dagger$ is a solution of the limiting dual problem* (5.7).

Proof. The source condition (5.8) guarantees the existence of a $\mu \in \mathcal{Y}$ s.t. $A^*\mu \in \partial\mathcal{J}(u^\dagger_{\mathcal{J}})$, and recall from Remark 5.1 that this guarantees the existence of a solution μ_0 to the limiting dual problem 5.7. Let us assume that μ_0 is an arbitrary solution to the limiting dual problem. Since we have $A^*\mu_\delta \in \partial\mathcal{J}(u_\delta) \subseteq \partial\mathcal{J}(0)$, μ_δ is feasible for this limiting dual problem, and we have

$$\langle \mu_\delta, f\rangle \leqslant \langle \mu_0, f\rangle, \quad \text{or} \quad 0 \leqslant \langle \mu_0 - \mu_\delta, f\rangle, \tag{5.10}$$

for any $\delta > 0$. Similarly, since μ_δ solves the dual problem (5.2) and μ_0 is feasible in (5.2), we get that for all δ

$$\langle \mu_0, f_\delta\rangle - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{Y}^*} \leqslant \langle \mu_\delta, f_\delta\rangle - \frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{Y}^*}, \quad \text{or} \quad \frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{Y}^*} - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{Y}^*} \leqslant \langle \mu_0 - \mu_\delta, -f_\delta\rangle. \tag{5.11}$$

Summing up these estimates, we get that

$$\frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{Y}^*} - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{Y}^*} \leqslant \langle \mu_0 - \mu_\delta, f - f_\delta\rangle \leqslant \delta\|\mu_0 - \mu_\delta\|_{\mathcal{Y}^*}.$$

Noting that

$$\frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{Y}^*} - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{Y}^*} = \frac{\alpha}{2}(\|\mu_\delta\|_{\mathcal{Y}^*} - \|\mu_0\|_{\mathcal{Y}^*})(\|\mu_\delta\|_{\mathcal{Y}^*} + \|\mu_0\|_{\mathcal{Y}^*}),$$

we get that

$$\frac{\alpha}{2}(\|\mu_\delta\|_{\mathcal{Y}^*} - \|\mu_0\|_{\mathcal{Y}^*})(\|\mu_0\|_{\mathcal{Y}^*} + \|\mu_\delta\|_{\mathcal{Y}^*}) \leqslant \delta\|\mu_0 - \mu_\delta\|_{\mathcal{Y}^*} \leqslant \delta(\|\mu_0\|_{\mathcal{Y}^*} + \|\mu_\delta\|_{\mathcal{Y}^*})$$

and

$$\|\mu_\delta\|_{\mathcal{Y}^*} \leqslant \|\mu_0\|_{\mathcal{Y}^*} + \frac{2\delta}{\alpha} \leqslant C, \tag{5.12}$$

since $\frac{\delta}{\alpha}$ is bounded. Since $\mathcal{Y}$ is assumed to be separable, the Banach–Alaoglu theorem (Theorem 4.1) tells us that for any sequence $\delta_n \downarrow 0$ there exists a subsequence (which we do not relabel) such that

$$\mu_{\delta_n} \rightharpoonup^* \mu^\dagger.$$

By weak-*-weak-* continuity of A^* and weak closedness of $\partial\mathcal{J}(0)$ (Theorem 4.28) we get that

$$A^*\mu^\dagger \in \partial\mathcal{J}(0)$$

and $\mu^\dagger$ is feasible for the limiting dual problem.

From (5.11) we obtain that

$$\begin{aligned}
\langle \mu_0, f_\delta\rangle &\leqslant \langle \mu_\delta, f_\delta\rangle + \frac{\alpha}{2}(\|\mu_0\|^2_{\mathcal{Y}^*} - \|\mu_{\delta_n}\|^2_{\mathcal{Y}^*}) \\
&\leqslant \langle \mu_\delta, f\rangle + \langle \mu_\delta, f_\delta - f\rangle + \frac{\alpha}{2}(\|\mu_0\|^2_{\mathcal{Y}^*} - \|\mu_{\delta_n}\|^2_{\mathcal{Y}^*}) \\
&\leqslant \langle \mu_\delta, f\rangle + \delta\|\mu_\delta\|_{\mathcal{Y}^*} + \frac{\alpha}{2}(\|\mu_0\|^2_{\mathcal{Y}^*} - \|\mu_{\delta_n}\|^2_{\mathcal{Y}^*}).
\end{aligned}$$

Taking the limit $\delta \to 0$, we get that

$$\langle \mu_0, f\rangle \leqslant \langle \mu^\dagger, f\rangle$$

Since $\mu^\dagger$ is feasible for the limiting dual problem and μ_0 is a solution of this problem, we conclude that $\mu^\dagger$ solves the limiting dual problem (5.7). □

The next theorem shows that the source condition (5.8) implies a convergence rate in terms of the Bregman divergence.

Theorem 5.7. *Let the source condition* (5.8) *be satisfied at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ and let u_δ be a regularised solution solving* (5.1)*. Then the following estimate holds*

$$D^{p_\delta,p^\dagger}_{\mathcal{J}}(u_\delta,u^\dagger_{\mathcal{J}}) \leqslant \frac{1}{4\alpha}\left(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*}\right)^2+\delta\|\mu^\dagger\|_{\mathcal{Y}^*}.$$

where $p_\delta = A^*\mu_\delta \in \partial\mathcal{J}(u_\delta)$ with μ_δ as defined in (5.3) *and $p^\dagger = A^*\mu^\dagger \in \partial\mathcal{J}(u^\dagger_{\mathcal{J}})$ is as defined in* (5.8)*. $D^{p_\delta,p^\dagger}_{\mathcal{J}}(u_\delta,u^\dagger_{\mathcal{J}})$ denotes the symmetric Bregman divergence between u_δ and $u^\dagger_{\mathcal{J}}$. For the optimal choice $\alpha=\delta/\|\mu^\dagger\|_{\mathcal{Y}^*}$ we get that*

$$D^{p_\delta,p^\dagger}_{\mathcal{J}}(u_\delta,u^\dagger_{\mathcal{J}}) \leqslant 3\delta\|\mu^\dagger\|_{\mathcal{Y}^*}.$$

Proof. We start with the following estimate

$$\begin{aligned}\alpha D^{p_\delta,p^\dagger}_{\mathcal{J}}(u_\delta,u^\dagger_{\mathcal{J}}) &= \alpha\langle p_\delta-p^\dagger,u_\delta-u^\dagger_{\mathcal{J}}\rangle\\ &= \alpha\langle\mu_\delta-\mu^\dagger,Au_\delta-f\rangle\\ &= \alpha\langle\mu_\delta,Au_\delta-f_\delta\rangle+\alpha\langle\mu_\delta,f_\delta-f\rangle-\alpha\langle\mu^\dagger,Au_\delta-f_\delta\rangle-\alpha\langle\mu^\dagger,f_\delta-f\rangle.\end{aligned}$$

From (5.5) we know that

$$\alpha\langle\mu_\delta,Au_\delta-f_\delta\rangle \leqslant -\frac{1}{2}\|Au_\delta-f_\delta\|^2_{\mathcal{Y}}.$$

and from (5.4) that $\alpha\|\mu_\delta\|_{\mathcal{Y}^*}=\|Au_\delta-f_\delta\|_{\mathcal{Y}}$. Using these estimates, the "Cauchy–Schwarz inequality" and the estimate $\|f-f_\delta\|_{\mathcal{Y}}\leqslant\delta$, we get

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

The right-hand side is the following negative-definite quadratic function of the scalar variable $\|Au_\delta-f_\delta\|_{\mathcal{Y}}$:

$$\varphi(t) := -\frac{1}{2}t^2+(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*})t+\alpha\delta\|\mu^\dagger\|_{\mathcal{Y}^*},\quad t\in\mathbf{R}_{\geqslant 0}.$$

It achieves its maximum at $t_0=(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*})$ and this maximum value is equal to

$$\varphi(t_0)=\frac{(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*})^2}{2}+\alpha\delta\|\mu^\dagger\|_{\mathcal{Y}^*}.$$

Substituting this into the above estimate for the Bregman divergence and dividing both sides by α, we get the desired estimate

$$D^{p_\delta,p^\dagger}_{\mathcal{J}}(u_\delta,u^\dagger_{\mathcal{J}}) \leqslant \frac{(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*})^2}{2\alpha}+\delta\|\mu^\dagger\|_{\mathcal{Y}^*}.$$

Differentiating the right-hand side w.r.t. α and setting the derivative to zero, we obtain the following optimality condition for α

$$0=\frac{2\alpha\|\mu^\dagger\|_{\mathcal{Y}^*}(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*})-(\delta+\alpha\|\mu^\dagger\|_{\mathcal{Y}^*})^2}{2\alpha^2}=\frac{\alpha^2\|\mu^\dagger\|^2_{\mathcal{Y}^*}-\delta^2}{2\alpha^2}$$

and

$$\alpha=\frac{\delta}{\|\mu^\dagger\|_{\mathcal{Y}^*}}.$$

With this optimal choice of α we get the following estimate

$$D_{\mathcal{J}}^{p_\delta, p^\dagger}(u_\delta, u_{\mathcal{J}}^\dagger) \leqslant 3\delta \|\mu^\dagger\|_{\mathcal{Y}^*}.$$

□

Remark 5.8. Of course, we do not know $\mu^\dagger$ since we don't know the $\mathcal{J}$-minimising solution $u_{\mathcal{J}}^\dagger$, but the theorem gives an optimal *rate* $\alpha \sim \delta$ for a priori parameter choice rules and a corresponding error estimate $D_{\mathcal{J}}^{p_\delta, p^\dagger}(u_\delta, u_{\mathcal{J}}^\dagger) = O(\delta)$.

Now we will look at two examples involving Total Variation to get a feeling for what the source condition "means".

Example 5.9 (Total Variation)**.** Let $\Omega \subset \mathbf{R}^2$ be a bounded domain with a C^∞ boundary. Let $\mathcal{X} = \mathrm{BV}(\Omega)$ and $\mathcal{Y} = L^2(\Omega)$ and $\mathcal{J}(\cdot) = \mathrm{TV}(\cdot)$. Recall the ROF problem

$$\min_{u \in \mathrm{BV}} \frac{1}{2}\|Iu - f_\delta\|_{L^2}^2 + \alpha\,\mathrm{TV}(u),$$

where $I\colon \mathrm{BV}(\Omega) \to L^2(\Omega)$ is the embedding operator, which is continuous since $\Omega \subset \mathbf{R}^2$. The adjoint $I^*\colon L^2(\Omega) \to \mathrm{BV}^*(\Omega)$ continuously embeds L^2 into BV^*. Clearly, I^* is not surjective and $\mathrm{im}(I^*) = L^2(\Omega)$.

From Example 4.58 we know that

$$\mathrm{TV}(\mathbf{1}_C) = \mathrm{Per}(C),$$

where $\mathbf{1}_C$ is the indicator function of the set C. Denoting by $\mathbf{n}_{\partial C}$ the unit normal, we obtain

$$\mathrm{Per}(C) = \int_{\partial C} 1 = \int_{\partial C} \langle \mathbf{n}_{\partial C}, \mathbf{n}_{\partial C} \rangle.$$

Since $\mathbf{n}_{\partial C} \in C^\infty(\partial C, \mathbf{R}^2)$ and $\|\mathbf{n}_{\partial C}(x)\|_2 = 1$ for any x, we can extend $\mathbf{n}_{\partial C}$ to a $C_0^\infty(\Omega, \mathbf{R}^2)$ vector field ψ with $\sup_{x\in\Omega} \|\psi(x)\|_2 \leqslant 1$. Therefore, using the divergence theorem, we obtain that

$$\int_{\partial C} \langle \mathbf{n}_{\partial C}, \mathbf{n}_{\partial C} \rangle = \int_{\partial C} \langle \psi, \mathbf{n}_{\partial C} \rangle = \int_C \operatorname{div} \psi = \int_\Omega \mathbf{1}_C \operatorname{div} \psi.$$

Combining all these equalities, we get that

$$\mathrm{TV}(\mathbf{1}_C) = \int_\Omega \mathbf{1}_C \operatorname{div} \psi = \langle \operatorname{div} \psi, \mathbf{1}_C \rangle.$$

Taking an arbitrary $u \in \mathrm{BV}(\Omega)$, we note that

$$\mathrm{TV}(u) - \langle \operatorname{div} \psi, u \rangle = \sup_{\varphi \in \mathcal{D}(\Omega, \mathbf{R}^n)} \langle \operatorname{div} \varphi, u \rangle - \langle \operatorname{div} \psi, u \rangle \geqslant 0,$$

since $\varphi = \psi$ is feasible. Therefore, $\operatorname{div} \psi \in \partial\,\mathrm{TV}(0)$ and, since $\mathrm{TV}(\mathbf{1}_C) = \langle \operatorname{div} \psi, \mathbf{1}_C \rangle$, we also get that

$$\operatorname{div} \psi \in \partial\,\mathrm{TV}(\mathbf{1}_C).$$

Since $\psi \in C_0^\infty(\Omega, \mathbf{R}^2)$, we have $\operatorname{div} \psi \in C_0^\infty(\Omega) \subset L^2(\Omega) = \mathrm{im}(I^*)$ and the source condition is satisfied at $u = \mathbf{1}_C$ with $\mu^\dagger = \operatorname{div} \psi$.

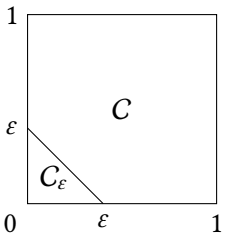

Figure 5.1: Example of a set whose indicator function does not satisfy the source condition.

Example 5.10 (Total Variation)**.** In the same setting as in Example 5.9, let C be a domain with a nonsmooth boundary, e.g., a square $C = [0,1]^2$. We will show in this example that in this case $\partial \mathrm{TV}(\mathbf{1}_C) \cap \mathrm{im}(I^*) = \emptyset$, where $\mathrm{im}(I^*) = L^2(\Omega)$ as before, i.e. the source condition fails.

Assume that there exists $p_0 \in \partial \mathrm{TV}(\mathbf{1}_C) \cap L^2(\Omega)$. Then by the results of Example 4.58 we have that

$$\langle p_0, \mathbf{1}_C \rangle = \mathrm{TV}(\mathbf{1}_C) = \mathrm{Per}(C) = 4.$$

Since p_0 is a subgradient, we get that for any $u \in \mathrm{BV}(\Omega)$

$$\mathrm{TV}(u) - \langle p_0, u \rangle \geqslant 0.$$

Let us cut a triangle C_ε of size ε from a corner of C as shown in Figure 5.1. Then for $u = \mathbf{1}_{C \setminus C_\varepsilon}$ we get

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

and therefore

$$\langle p_0, \mathbf{1}_{C_\varepsilon} \rangle \geqslant \mathrm{TV}(\mathbf{1}_C) - \mathrm{TV}(\mathbf{1}_{C \setminus C_\varepsilon}) = \mathrm{Per}(C) - \mathrm{Per}(C \setminus C_\varepsilon) = 4 - (4 - 2\varepsilon + \sqrt{2}\varepsilon) = (2 - \sqrt{2})\varepsilon > 0.$$

By Hölder's inequality we get that

$$\langle p_0, \mathbf{1}_{C_\varepsilon} \rangle = \int_{C_\varepsilon} p_0 \cdot \mathbf{1} \leqslant \left(\int_{C_\varepsilon} |p_0|^2 \right)^{1/2} \left(\int_{C_\varepsilon} 1 \right)^{1/2} = \frac{1}{\sqrt{2}} \varepsilon \left(\int_{C_\varepsilon} |p_0|^2 \right)^{1/2}.$$

Combining the last two inequalities, we get

$$(2 - \sqrt{2})\varepsilon \leqslant \langle p_0, \mathbf{1}_{C_\varepsilon} \rangle \leqslant \frac{1}{\sqrt{2}} \varepsilon \left(\int_{C_\varepsilon} |p_0|^2 \right)^{1/2}$$

and therefore

$$\int_{C_\varepsilon} |p_0|^2 \geqslant 2(2 - \sqrt{2})^2 > 0$$

for all $\varepsilon > 0$. However, since $p_0 \in L^2(\Omega)$ by assumption, we must have

$$\int_{C_\varepsilon} |p_0|^2 \to 0 \quad \text{as } \varepsilon \to 0.$$

This contradiction proves that such p_0 does not exist and $\partial \mathrm{TV}(\mathbf{1}_C) \cap \mathrm{im}(I^*) = \emptyset$.

Appendix A

Sobolev spaces

Sobolev spaces constitute one of the most relevant functional settings for the treatment of PDEs and boundary value problems. This appendix gives a short introduction to the topic. Sobolev spaces are covered in more detail in the course *Analysis of Partial Differential Equations*. For further background, including applications to PDEs, see [13].

We start by introducing the notion of a weak derivatives that generalises the classical partial derivatives.

Definition A.1 (Test functions)**.** *Let $O \subseteq \mathbf{R}^d$. We set*

$$C_0^\infty(O) = \{\varphi \in C^\infty(O) | \operatorname{supp}(\varphi) \Subset O\},$$

the smooth functions with compact support. This space is often referred as the space of test functions and denoted by $\mathcal{D}(O)$.

If $u \in C^1(\mathbf{R})$ then we can define $\frac{\partial u}{\partial x}$ by

$$\int \frac{\partial u}{\partial x}(x)\varphi(x)\,\mathrm{d}x = -\int u(x)\frac{\partial \varphi}{\partial x}(x)\,\mathrm{d}x,$$

for all $\varphi \in \mathcal{D}(\mathbf{R})$. We notice that the right hand side is well-defined for all $u \in L^1_{\mathrm{loc}}(\mathbf{R})$.

Definition A.2. *Let $\alpha = \alpha_1, \cdots, \alpha_d$ be a multi-index, $\alpha_i \in \mathbf{N}$, and $|\alpha| = \alpha_1 + \cdots + \alpha_d$. A function $u \in L^1_{\mathrm{loc}}(O)$ has a weak derivative $v = D^\alpha u = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d} u \in L^1_{\mathrm{loc}}(O)$ if*

$$\int_O v(x)\varphi(x)\,\mathrm{d}x = (-1)^{|\alpha|}\int_O u(x) D^\alpha \varphi(x)\,\mathrm{d}x,$$

for all test functions $\varphi \in \mathcal{D}(O)$.

Note that when the weak derivative $D^\alpha u$ exists, it is defined only up to a set of measure zero. So any point-wise statements to be made about $D^\alpha u$ are understood to only hold almost surely. Most of the classical differential calculus can be reproduced for weak derivatives (e.g. the product rule and the chain rule).

Definition A.3. *The Sobolev space $H^s(O)$, $s \in \mathbf{N}$, is defined as the set of all functions $u \in L^2(O)$ with weak derivatives $D^\alpha u \in L^2(O)$ up to the order $|\alpha| \leqslant s$.*

The above definition can be generalised for functions $u \in L^p(O)$, $1 \leqslant p \leqslant \infty$, and the resulting Sobolev spaces are usually denoted by $W^{s,p}(O)$. The Sobolev spaces $H^s(O)$ are Banach spaces with the norm

$$\|u\|_{H^s} = \left(\sum_{|\alpha| \leqslant s} \int_O \|D^\alpha u\|^2_{L^2(O)} \, \mathrm{d}x \right)^{\frac{1}{2}}, \tag{A.1}$$

and, in fact, this norm is induced by the following inner product, so these spaces are Hilbert spaces:

$$\langle u, v \rangle_{H^s} = \sum_{|\alpha| \leqslant s} \langle D^\alpha u, D^\alpha v \rangle_{L^2} = \sum_{|\alpha| \leqslant s} \int_O D^\alpha u(x) D^\alpha v(x) \, \mathrm{d}x,$$

for all $u, v \in H^s(O)$.

Bibliography

[1] Y. A. Abramovich and C. D. Aliprantis, *An Invitation to Operator Theory*, Graduate Studies in Mathematics, American Mathematical Society, 2002.

[2] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Elsevier Science, Singapore, 2003.

[3] C. D. Aliprantis and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 2006.

[4] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Clarendon Press, 2000.

[5] A. B. Bakushinskii, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, Zhurnal Vychislitel'noï Matematiki i Matematicheskoï Fiziki, 24 (1984), pp. 1258–1259.

[6] M. Benning and M. Burger, *Modern regularization methods for inverse problems*, Acta Numerica, 27 (2018), pp. 1–111.

[7] B. Bollobás, *Linear Analysis: An Introductory Course*, Cambridge University Press, Cambridge, second ed., 1999.

[8] K. Bredies and D. A. Lorenz, *Mathematical Image Processing*, Springer, 2018.

[9] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), p. 1411.

[10] ———, *A guide to the tv zoo*, in Level-Set and PDE-based Reconstruction Methods, M. Burger and S. Osher, eds., Springer, 2013.

[11] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, 1976.

[12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375, Springer Science & Business Media, 1996.

[13] L. C. Evans, *Partial differential equations*, American Mathematical Society, Providence, RI, 1998.

[14] J. Hadamard, *Sur les problèmes aux dérivés partielles et leur signification physique*, Princeton University Bulletin, 13 (1902), pp. 49–52.

[15] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*, Springer Science & Business Media, 2000.

[16] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.

[17] W. Rudin, *Functional Analysis*, International series in pure and applied mathematics, McGraw-Hill, 1991.

[18] B. P. Rynne and M. A. Youngson, *Linear Functional Analysis*, Springer Undergraduate Mathematics Series, Springer, London, 2nd ed ed., 2008.

[19] K. Saxe, *Beginning Functional Analysis*, Springer, 2002.

[20] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Springer, 2009.

[21] T. Tao, *Epsilon of Room, One*, vol. 1, American Mathematical Soc., 2010.

[22] E. Zeidler, *Applied Functional Analysis: Applications to Mathematical Physics*, vol. 108 of Applied Mathematical Sciences Series, Springer, 1995.

[23] ——, *Applied Functional Analysis: Main Principles and Their Applications*, vol. 109 of Applied Mathematical Sciences Series, Springer, 1995.