Inverse Problems in Imaging

Yury Korolev

Last updated on: November 27, 2018

Lecture Notes
Michelmas Term 2018/2019

Contents

These lecture notes use the material in the notes of the course "Inverse Problems", which was held by Matthias J. Ehrhardt and Lukas Lang in Lent term 2018 at the University of Cambridge.[1] Complementary material can be found in the following books, lecture notes and review papers:

(a) Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems.* Vol. 375. Springer Science & Business Media, 1996.

(b) Martin Burger. *Inverse Problems.* Lecture notes winter semester 2007/2008.

`http://www.math.uni-muenster.de/num/Vorlesungen/IP_WS07/skript.pdf`

(c) Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier and Frank Lenzen. *Variational Methods in Imaging.* Applied Mathematical Sciences, Springer New York, 2008.

(d) Kristian Bredies and Dirk Lorenz. *Mathematische Bildverarbeitung (in German).* Vieweg+Teubner Verlag, Springer Fachmedien Wiesbaden, 2011

(e) Martin Burger and Stanley Osher. *A guide to the TV Zoo.* In: Level set and PDE based reconstruction methods in imaging. Springer, 2013.

`https://www.springer.com/cda/content/document/cda_downloaddocument/9783319017112-c1.pdf?SGWID=0-0-45-1422511-p175374869`

(f) Antonine Chambolle and Thomas Pock. *An introduction to continuous optimization for imaging.* Acta Numerica, 25, 161-319 (2016)

`https://www.cambridge.org/core/journals/acta-numerica/article/an-introduction-to-continuous-optimization-for-imaging/1115AA7E36FC201E811040D11118F67F`

(g) Andreas Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems.* Vol. 120. Springer Science & Business Media, 1996.

(h) Kazufumi Ito and Bangti Jin. *Inverse Problems: Tikhonov Theory and Algorithms.* World Scientific, 2014.

(i) Jennifer L. Mueller and Samuli Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications.* Vol. 10. SIAM, 2012.

These lecture notes are under constant redevelopment and might contain typos or errors. I very much appreciate the finding and reporting of those (to y.korolev@damtp.cam.ac.uk). Thanks!

[1] `http://www.damtp.cam.ac.uk/research/cia/teaching/201718lentinvprob.html`

Chapter 1

Introduction to Inverse Problems

Inverse problems are usually concerned with the interpretation of indirect measurements. They assume that there is a connection between the quantities of interest and measured data that is referred to as the *forward model* or the *forward operator*. The forward operator typically models the physics of data acquisition.

Inverse problems arise in many different fields of science and technology, such as medical imaging, geoscience, climate studies, astronomy, microscopy, non-destructive testing and many others. Mathematically, they are formulated as operator equations

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

where $A\colon \mathcal{U} \to \mathcal{V}$ is the forward operator acting between some spaces $\mathcal{U}$ and $\mathcal{V}$, typically Hilbert or Banach spaces, f are the measured data and u is the quantity we want to reconstruct from the data.

Hadamard [16] gave a definition that was supposed to describe the class of mathematical problems that could be reasonably solved in practice. Applied to problem (1.1), it reads as follows.

Definition 1.0.1. *The problem* (1.1) *is called well-posed if*

- *it has a solution* $\forall f \in \mathcal{V}$*,*
- *the solution is unique,*
- *the solution depends continuously on the data, i.e. small errors in the data* f *result in small errors in the reconstruction.*

If any of these properties are not satisfied, the problem is called ill-posed.

Some of the notions in this definition, like the notions of "small", need to be made precise, for example, using norms in particular spaces. This choice will affect the well- or ill-posedness of the problem. It turns out, however, that under reasonable (and realistic) choices of norms many inverse problems are ill-posed. Their practical significance lead to the development of the so-called *regularisation theory* that provides a mathematical basis for dealing with ill-posed problems.

In this course we will develop this theory for *linear* inverse problems (i.e. those,w here the forward operator A is a linear bounded operator). This class includes such important applications as computer tomography, magnetic resonance imaging and image deblurring in microscopy or astronomy. There are, however, many other important applications, such as seismic imaging, where the forward operator in non-linear (e.g., parameter identification problems for PDEs).

1.1 Examples of Inverse Problems

1.1.1 Differentiation

Consider the problems of evaluation 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 1.1.1. *The operator D is unbounded from $L^2[0, \pi/2] \to L^2[0, \pi/2]$.*

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

This shows that differentiation is ill-posed 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 L^2) to L^2. Indeed, $\forall u \in H^1$ we get

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

However, since in practice we typically deal with functions corrupted by nonsmooth noise, the L^2 setting is practice-relevant, 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

$$f(x) = \int_0^x u(t)\, dt,$$

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

1.1.2 Image Deblurring and Denoising

Whenever a camera (or a microscope, or a telescope) records an image, what is recorded is an integral transform of the image:

$$f(x) = (Au)(x) := \int K(x, \xi) u(\xi)\, d\xi, \tag{1.2}$$

Here $u(\xi)$ is the 'true' image and $K(x, \xi)$ is the so-called point-spread function (PSF). The PSF models the optics of the camera. If a bright source localised at $\xi = \xi_0$ is recorded (i.e., $u(\xi)$ can be well approximated by the Dirac delta $\delta(\xi - \xi_0)$), the result has the form $f(x) = K(x, \xi_0)$. This function is typically not localised, i.e. the point source gets spread out, hence the name of the PSF. For non-localised images, the effect is that the image gets blurred. The task of restoring a sharp image from its blurred version is called *image deblurring* and plays a crucial role in the processing of microscopy and astronomy images. Apart from blur, recorded images often suffer from measurement noise; therefore, the task of image deblurring is often complemented with that of *denoising.*

A special case of problem (1.2) is when the PSF is spatially invariant and takes the form

$$K(x, \xi) = K(x - \xi).$$

In this case, the integral transform (1.2) is a *convolution* and the reconstruction problem is referred to as *deconvolution.*

Theorem 1.1.2 (e.g., [12, Thm. XI.6.6]). *Let $A\colon L^2(\Omega) \to L^2(\Omega)$ be as defined in* (1.2) *with $K(\cdot,\cdot) \in L^2(\Omega \times \Omega)$ (in this case A is called a Hilbert-Schmidt operator). Then A is compact.*

Remark 1.1.3. As we shall see later in the course, inversion of a compact operator is ill-posed and, therefore, deblurring (and its special case, deconvolution) is an ill-posed problem.

Remark 1.1.4. In some cases one considers the task of denoising separately, i.e. one considers the problem (1.1) with A be the identity operator. This is not really an inverse problem, but it is usually solved using the same methods as problem (1.1).

1.1.3 Matrix Inversion

In finite dimensions, the inverse problem (1.1) becomes 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 $u \in \mathbb{R}^n$ and $f \in \mathbb{R}^n$ being n-dimensional vectors with real entries and $A \in \mathbb{R}^{n\times n}$ being a matrix with real entries. Assume further A to be 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 $a_j \in \mathbb{R}^n$ for $j \in \{1,\ldots,n\}$ such that A can be written as

$$A = \sum_{j=1}^{n} \lambda_j a_j a_j^\top . \tag{1.3}$$

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

We assume that we measure f_δ instead of f, with $\|f - f_\delta\|_2 \leqslant \delta\|A\| = \delta\lambda_1$, where $\|\cdot\|_2$ denotes the Euclidean norm of $\mathbb{R}^n$ 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*. We want to demonstrate the effect of this error amplification with a small example.

Example 1.1.5. Let us 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 $\|u - u_\delta\|/\delta \approx 4000 \approx \kappa$.

1.1.4 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 $u \in C_0^\infty(\mathbb{R}^n)$, $s \in \mathbb{R}$, and $\theta \in S^{n-1}$ the *Radon transform* $R : C_0^\infty(\mathbb{R}^n) \to C^\infty(S^{n-1} \times \mathbb{R})$ can be defined as the integral operator

$$\begin{aligned} f(\theta, s) = (\mathcal{R}u)(\theta, s) &= \int_{x\cdot\theta = s} u(x)\, dx \\ &= \int_{\theta^\perp} u(s\theta + y)\, dy, \end{aligned} \tag{1.4}$$

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

$$f(\theta, s) = (\mathcal{P}u)(\theta, s) = \int_{\mathbb{R}} u(s\theta + t\theta^\perp)\, dt,$$

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 $\mathbb{R}^n$, see Fig. 1.1[2].

Example 1.1.6. Let $n = 2$. Then S^{n-1} is simply the unit sphere $S^1 = \{\theta \in \mathbb{R}^2 \mid \|\theta\| = 1\}$. 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.

$$f(\varphi, s) = (\mathcal{R}u)(\varphi, s) = \int_{\mathbb{R}} u(s\cos(\varphi) - t\sin(\varphi), s\sin(\varphi) + t\cos(\varphi))\, dt. \tag{1.5}$$

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.

Remark 1.1.7 ([20, p. 38]). 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.

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

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

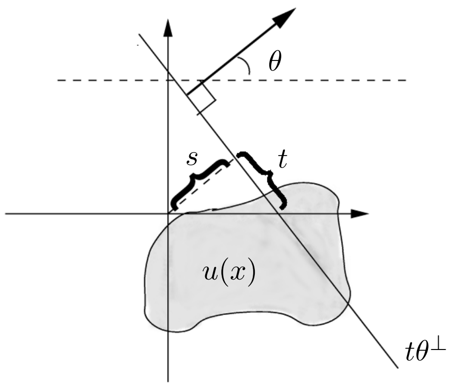

Figure 1.1: Visualization of the Radon transform in two dimensions (which coincides with the X-ray transform). The function u is integrated over the ray parametrized by θ and s.[3]

X-ray Computed Tomography (CT)

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.

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

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

$$\frac{d}{dt} I(x + t\theta) = -I(x + t\theta)u(x + t\theta), \tag{1.6}$$

Let $R > 0$ be the radius of the domain of interest centred at the origin. Then, we integrate (1.6) 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

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

Note that, due to $d/dx \log(f(x)) = f'(x)/f(x)$, the left hand side in the above equation simplifies to

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

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 (1.5) (if we further assume that u can be continuously extended to zero outside of the circle of radius R).

Example 1.1.8 ([20, p. 38])**.** Let $B := \{x \in \mathbb{R}^2 \mid \|x\| \leqslant 1\}$ denote the unit ball in $\mathbb{R}^2$ and $Z := [-1, 1] \times [0, \pi)$. Moreover, let $\theta(\varphi) := (\cos(\varphi), \sin(\varphi))^\top$, $\theta^\perp(\varphi) := (\sin(\varphi), -\cos(\varphi))^\top$

be the unit vectors pointing in the direction described by φ and orthogonal to it. Then, the Radon transform/X-ray transform is defined as the operator $R\colon L^2(B) \to L^2(Z)$ with

$$(Ru)(s,\varphi) := \int_{-\sqrt{1-s^2}}^{\sqrt{1-s^2}} u\Big(s\theta(\varphi) + t\theta^{\perp}(\varphi)\Big)\, dt.$$

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, i.e. $R \in \mathcal{K}(L^2(B), L^2(Z))$.

Positron Emission Tomography (PET)

In Positron Emission Tomography (PET) a so-called radioactive tracer (a positron emitting radionuclide on a biologically active molecule) is injected into a patient (or subject). The emitted positrons of the tracer will interact with the subjects' electrons after travelling a short distance (usually less than 1mm), causing the annihilation of both the positron and the electron, which results in a pair of gamma rays moving into (approximately) opposite directions. This pair of photons is detected by the scanner detectors, and an intensity $f(\varphi, s)$ can be associated with the number of annihilations detected at the detector pair that forms the line with offset s and angle φ (with respect to the reference coordinate system). Thus, we can consider the problem of recovering the unknown tracer density u as a solution of the inverse problem (1.4) again. The line of integration is determined by the position of the detector pairs and the geometry of the scanner.

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., [8, 25]. Nevertheless, we shall 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. $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ with

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

For $A \colon \mathcal{U} \to \mathcal{V}$ we further want to denote by

(a) $\mathcal{D}(A) := \mathcal{U}$ the domain,

(b) $\mathcal{N}(A) := \{u \in \mathcal{U} \mid Au = 0\}$ the kernel,

(c) $\mathcal{R}(A) := \{f \in \mathcal{V} \mid f = Au, u \in \mathcal{U}\}$ the range

of A.

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

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

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

$$\|Au\|_{\mathcal{V}} \leqslant C\|u\|_{\mathcal{U}}$$

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

In this Chapter we only consider $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ with $\mathcal{U}$ and $\mathcal{V}$ being Hilbert spaces. From functional calculus we know that every Hilbert space $\mathcal{U}$ is equipped with a *scalar 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 scalar product structure together with the theorem of Fréchet-Riesz [25, Section 2.10, Theorem 2.E] allows us to define the (unique) *adjoint operator* of A, denoted with A^*, as follows:

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

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

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

One can show that $\mathcal{X}^\perp$ is a closed subspace and that $\mathcal{U}^\perp = \{0\}$. 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 there exists the *orthogonal decomposition*

$$\mathcal{U} = \mathcal{X} \oplus \mathcal{X}^\perp,$$

which means that every element $u \in \mathcal{U}$ 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 [25, Section 2.9, Corollary 1].

The mapping $u \mapsto x$ defines a linear operator $P_\mathcal{X} \in \mathcal{L}(\mathcal{U}, \mathcal{U})$ that is called *orthogonal projection* on $\mathcal{X}$.

Lemma 2.0.1 (cf. [19, Section 5.16])**.** *Let $\mathcal{X} \subset \mathcal{U}$ be a closed subspace. The orthogonal projection onto $\mathcal{X}$ satisfies the following conditions:*

(a) $P_\mathcal{X}$ is self-adjoint, i.e. $P_\mathcal{X}^* = P_\mathcal{X}$,

(b) $\|P_\mathcal{X}\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} = 1$ (if $\mathcal{X} \neq \{0\}$),

(c) $I - P_\mathcal{X} = P_{\mathcal{X}^\perp}$,

(d) $\|u - P_\mathcal{X} u\|_\mathcal{U} \leqslant \|u - v\|_\mathcal{U}$ for all $v \in \mathcal{X}$,

(e) $x = P_\mathcal{X} u$ if and only if $x \in \mathcal{X}$ and $u - x \in \mathcal{X}^\perp$.

Remark 2.0.2. Note that for a non-closed subspace $\mathcal{X}$ we only have $(\mathcal{X}^\perp)^\perp = \overline{\mathcal{X}}$. For $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ we therefore have

- $\mathcal{R}(A)^\perp = \mathcal{N}(A^*)$ and thus $\mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$,
- $\mathcal{R}(A^*)^\perp = \mathcal{N}(A)$ and thus $\mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}$.

Hence, we can deduce the following orthogonal decompositions

$$\mathcal{U} = \mathcal{N}(A) \oplus \overline{\mathcal{R}(A^*)} \text{ and } \mathcal{V} = \mathcal{N}(A^*) \oplus \overline{\mathcal{R}(A)}.$$

We will also need the follwoing relationship between the ranges of A^* and A^*A.

Lemma 2.0.3. *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then $\overline{\mathcal{R}(A^*A)} = \overline{\mathcal{R}(A^*)}$.*

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

Let $u \in \overline{\mathcal{R}(A^*)}$ and let $\varepsilon > 0$. Then, there exists $f \in \mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$ with $\|A^*f - u\|_\mathcal{U} < \varepsilon/2$ (recall the orthogonal decomposition in Remark 2.0.2). As $\mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$, there exists $x \in \mathcal{U}$ such that $\|Ax - f\|_\mathcal{V} < \varepsilon/(2\|A\|_{\mathcal{L}(\mathcal{U},\mathcal{V})})$. Putting these together we have

$$\begin{aligned}\|A^*Ax - u\|_\mathcal{U} &\leqslant \|A^*Ax - A^*f\|_\mathcal{U} + \|A^*f - u\|_\mathcal{U} \\ &\leqslant \underbrace{\|A^*\|_{\mathcal{L}(\mathcal{U},\mathcal{V})}\|Ax - f\|_\mathcal{V}}_{<\varepsilon/2} + \underbrace{\|A^*f - u\|_\mathcal{U}}_{<\varepsilon/2} < \varepsilon\end{aligned}$$

which shows that $u \in \overline{\mathcal{R}(A^*A)}$ and thus also $\overline{\mathcal{R}(A^*)} \subseteq \overline{\mathcal{R}(A^*A)}$. $\square$

2.1 Generalised Inverses

Recall the inverse problem

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

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

Definition 2.1.1 (Minimal-norm solutions). *An element $u \in \mathcal{U}$ is called*

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

$$\|Au - f\|_{\mathcal{V}} = \inf\{\|Av - f\|_{\mathcal{V}}, \quad v \in \mathcal{U}\};$$

- *a minimal-norm solution of* (2.1) *(and is denoted by $u^\dagger$) if*

$$\|u^\dagger\|_{\mathcal{U}} \leqslant \|v\|_{\mathcal{U}} \quad \textit{for all least squares solutions } v.$$

Remark 2.1.2. Since $\mathcal{R}(A)$ 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 minimal-norm solution is unique (it is the orthogonal projection of the zero element onto an affine subspace defined by $\|Au - f\|_{\mathcal{V}} = \min\{\|Av - f\|_{\mathcal{V}}, \quad v \in \mathcal{U}\}$).

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

Theorem 2.1.3. *Let $f \in \mathcal{V}$ and $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then, the following three assertions are equivalent.*

(a) $u \in \mathcal{U}$ satisfies $Au = P_{\overline{\mathcal{R}(A)}} f$.

(b) u is a least squares solution of the inverse problem (2.1)*.*

(c) u solves the normal equation

$$A^*Au = A^*f. \tag{2.2}$$

Remark 2.1.4. The name normal equation is derived from the fact that for any solution u its residual $Au - f$ is orthogonal (normal) to $\mathcal{R}(A)$. This can be readily seen, as we have for any $v \in \mathcal{U}$ that

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

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

Proof of Theorem 2.1.3. For (a) $\Rightarrow$ (b): Let $u \in \mathcal{U}$ such that $Au = P_{\overline{\mathcal{R}(A)}} f$ and let $v \in \mathcal{U}$ be arbitrary. With the basic properties of the orthogonal projection, Lemma 2.0.1 (d), we have

$$\|Au - f\|_{\mathcal{V}}^2 = \|(I - P_{\overline{\mathcal{R}(A)}})f\|_{\mathcal{V}}^2 \leqslant \inf_{g \in \overline{\mathcal{R}(A)}} \|g - f\|_{\mathcal{V}}^2 \leqslant \inf_{v \in \mathcal{U}} \|Av - f\|_{\mathcal{V}}^2,$$

which shows that u is a least squares solution. Here, the last inequality follows from $\mathcal{R}(A) \subset \overline{\mathcal{R}(A)}$.

For (b) $\Rightarrow$ (c): Let $u \in \mathcal{U}$ be a least squares solution and let $v \in \mathcal{U}$ an arbitrary element. We define the quadratic polynomial $F\colon \mathbb{R} \to \mathbb{R}$,

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

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

For (c) $\Rightarrow$ (a): From the normal equation it follows that $A^*(f - Au) = 0$, which is equivalent to $f - Au \in \mathcal{R}(A)^\perp$, see Remark 2.1.4. Since $\mathcal{R}(A)^\perp = \left(\overline{\mathcal{R}(A)}\right)^\perp$ and $Au \in \mathcal{R}(A) \subset \overline{\mathcal{R}(A)}$, the assertion follows from Lemma 2.0.1 (e):

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

$\square$

Lemma 2.1.5. *Let $f \in \mathcal{V}$ and let $\mathbb{L}$ be the set of least squares solutions to the inverse problem* (2.1)*. Then, $\mathbb{L}$ is non-empty if and only if $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$.*

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

Consider now $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$. Then there exists $u \in \mathcal{U}$ and $g \in \mathcal{R}(A)^\perp = \left(\overline{\mathcal{R}(A)}\right)^\perp$ such that $f = Au + g$ and thus $P_{\overline{\mathcal{R}(A)}} f = P_{\overline{\mathcal{R}(A)}} Au + P_{\overline{\mathcal{R}(A)}} g = Au$ and the assertion follows from Theorem 2.1.3 (a). $\square$

Remark 2.1.6. If the dimensions of $\mathcal{U}$ and $\mathcal{R}(A)$ are finite, then $\mathcal{R}(A)$ is closed, i.e. $\overline{\mathcal{R}(A)} = \mathcal{R}(A)$. Thus, in a finite dimensional setting, there always exists a least squares solution.

Theorem 2.1.7. *Let $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$. Then there exists a unique minimal norm solution $u^\dagger$ to the inverse problem* (2.1) *and all least squares solutions are given by $\{u^\dagger\} + \mathcal{N}(A)$.*

Proof. From Lemma 2.1.5 we know that there exists a least squares solution. As noted in Remark 2.1.2, in this case the minimal-norm solution is unique. Let φ be an arbitrary least-squares solution. Using Theorem 2.1.3 we get

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

which shows that $\varphi - u^\dagger \in \mathcal{N}(A)$, hence the assertion. $\square$

If a least-squares solution exists for a given $f \in \mathcal{V}$ then the minimal-norm solution can be computed (at least in theory) using the Moore-Pensrose generalised inverse.

Definition 2.1.8. *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and let*

$$\tilde{A} := A|_{\mathcal{N}(A)^\perp} : \mathcal{N}(A)^\perp \to \mathcal{R}(A)$$

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

$$\mathcal{D}(A^\dagger) = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$$

with

$$\mathcal{N}(A^\dagger) = \mathcal{R}(A)^\perp .$$

Remark 2.1.9. Due to the restriction to $\mathcal{N}(A)^\perp$ and $\mathcal{R}(A)$ we have that $\tilde{A}$ is injective and surjective. Hence, $\tilde{A}^{-1}$ exists and is linear and – as a consequence – $A^\dagger$ is well-defined on $\mathcal{R}(A)$.

Moreover, due to the orthogonal decomposition $\mathcal{D}(A^\dagger) = \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$, there exist for arbitrary $f \in \mathcal{D}(A^\dagger)$ elements $f_1 \in \mathcal{R}(A)$ and $f_2 \in \mathcal{R}(A)^\perp$ with $f = f_1 + f_2$. Therefore, we have

$$A^\dagger f = A^\dagger f_1 + A^\dagger f_2 = A^\dagger f_1 = \tilde{A}^{-1} f_1 = \tilde{A}^{-1} P_{\overline{\mathcal{R}(A)}} f \,, \tag{2.4}$$

where we used that $f_2 \in \mathcal{R}(A)^\perp = \mathcal{N}(A^\dagger)$. Thus, $A^\dagger$ is well-defined on the entire domain $\mathcal{D}(A^\dagger)$.

Remark 2.1.10. As orthogonal complements are always closed we get that

$$\overline{\mathcal{D}(A^\dagger)} = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp = \mathcal{V},$$

and hence, $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$. Thus, if $\mathcal{R}(A)$ is closed it follows that $\mathcal{D}(A^\dagger) = \mathcal{V}$ and on the other hand, $\mathcal{D}(A^\dagger) = \mathcal{V}$ implies $\mathcal{R}(A)$ is closed. We note that for ill-posed problems $\mathcal{R}(A)$ is usually not closed; for instance, if A is compact then $\mathcal{R}(A)$ 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.

Theorem 2.1.11 ([14, Prop. 2.4])**.** *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then $A^\dagger$ is continuous, i.e. $A^\dagger \in \mathcal{L}(\mathcal{D}(A^\dagger), \mathcal{U})$, if and only if $\mathcal{R}(A)$ is closed.*

Example 2.1.12. To illustrate the definition of the Moore-Penrose inverse we consider a simple example in finite dimensions. Let the linear operator $A\colon \mathbb{R}^3 \to \mathbb{R}^2$ 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 $\mathcal{R}(A) = \{f \in \mathbb{R}^2 \mid f_2 = 0\}$ and $\mathcal{N}(A) = \{x \in \mathbb{R}^3 \mid x_1 = 0\}$. Thus, $\mathcal{N}(A)^\perp = \{x \in \mathbb{R}^3 \mid x_2, x_3 = 0\}$. Therefore, $\tilde{A}\colon \mathcal{N}(A)^\perp \to \mathcal{R}(A)$, given by $x \mapsto (2x_1, 0)^\top$, is bijective and its inverse $\tilde{A}^{-1}\colon \mathcal{R}(A) \to \mathcal{N}(A)^\perp$ is given by $f \mapsto (f_1/2, 0, 0)^\top$.

To get the Moore-Penrose inverse $A^\dagger$, we need to extend $\tilde{A}^{-1}$ to $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ in such a way that $A^\dagger f = 0$ for all $f \in \mathcal{R}(A)^\perp = \{f \in \mathbb{R}^2 \mid f_1 = 0\}$. It is easy to see that the Moore-Penrose inverse $A^\dagger\colon \mathbb{R}^2 \to \mathbb{R}^3$ 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 $\tilde{f} = (8,1)^\top \notin \mathcal{R}(A)$. Then, $A^\dagger \tilde{f} = A^\dagger (8,1)^\top = (4,0,0)^\top$.

It can be shown that $A^\dagger$ can be characterised by the Moore-Penrose equations.

Lemma 2.1.13 ([14, Prop. 2.3])**.** *The Moore-Penrose inverse $A^\dagger$ satisfies $\mathcal{R}(A^\dagger) = \mathcal{N}(A)^\perp$ and the Moore-Penrose equations*

(a) $AA^\dagger A = A$,

(b) $A^\dagger A A^\dagger = A^\dagger$,

(c) $A^\dagger A = I - P_{\mathcal{N}(A)}$,

(d) $AA^\dagger = P_{\overline{\mathcal{R}(A)}}\Big|_{\mathcal{D}(A^\dagger)}$,

where $P_{\mathcal{N}(A)}$ and $P_{\overline{\mathcal{R}(A)}}$ denote the orthogonal projections on $\mathcal{N}(A)$ and $\overline{\mathcal{R}(A)}$, respectively.

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

Theorem 2.1.14. *For each $f \in \mathcal{D}(A^\dagger)$, the minimal norm solution $u^\dagger$ to the inverse problem (2.1) is given via*

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

Proof. As $f \in \mathcal{D}(A^\dagger)$, we know from Theorem 2.1.7 that the minimal norm solution $u^\dagger$ exists and is unique. With $u^\dagger \in \mathcal{N}(A)^\perp$, Lemma 2.1.13, and Theorem 2.1.3 we conclude that

$$u^\dagger = (I - P_{\mathcal{N}(A)})u^\dagger = A^\dagger A u^\dagger = A^\dagger P_{\overline{\mathcal{R}(A)}} f = A^\dagger A A^\dagger f = A^\dagger f.$$

□

As a consequence of Theorem 2.1.14 and Theorem 2.1.3, 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 2.2.1. *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then A is said to be* compact *if for any bounded set $B \subset \mathcal{U}$ the closure of its image $\overline{A(B)}$ is compact in $\mathcal{V}$. We denote the space of compact operators by $\mathcal{K}(\mathcal{U}, \mathcal{V})$.*

Remark 2.2.2. We can equivalently define an operator A to be compact if the image of a bounded sequence $\{u_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$ contains a convergent subsequence $\{Au_{j_k}\}_{k\in\mathbb{N}} \subset \mathcal{V}$.

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 if ill-posedness.

Theorem 2.2.3. *Let $A \in \mathcal{K}(\mathcal{U}, \mathcal{V})$ with an infinite dimensional range. Then, the Moore-Penrose inverse of A is discontinuous.*

Proof. As the range $\mathcal{R}(A)$ is of infinite dimension, we can conclude that $\mathcal{U}$ and $\mathcal{N}(A)^\perp$ are also infinite dimensional. We can therefore find a sequence $\{u_j\}_{j\in\mathbb{N}}$ with $u_j \in \mathcal{N}(A)^\perp$, $\|u_j\|_{\mathcal{U}} = 1$ and $\langle u_j, u_k\rangle_{\mathcal{U}} = 0$ 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 $\|f_j - f_k\|_{\mathcal{V}} < \delta$. However, we also obtain

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

which shows that $A^\dagger$ is discontinuous. Here, the second identity follows from Lemma 2.1.13 (c) and the fact that $u_j, u_k \in \mathcal{N}(A)^\perp$. □

To have a better understanding of when we have $f \in \overline{\mathcal{R}(A)} \backslash \mathcal{R}(A)$ for compact operators A, we want to consider the singular value decomposition of compact operators.

Singular value decomposition of compact operators

Theorem 2.2.4 ([17, p. 225, Theorem 9.16])**.** *Let $\mathcal{U}$ be a Hilbert space and $A \in \mathcal{K}(\mathcal{U}, \mathcal{U})$ be self-adjoint. Then there exists an orthonormal basis $\{x_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$ of $\overline{\mathcal{R}(A)}$ and a sequence of eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}} \subset \mathbb{R}$ with $|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots > 0$ such that for all $u \in \mathcal{U}$ we have*

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

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

Remark 2.2.5. The notation in the theorem above only makes sense if the sequence $\{\lambda_j\}_{j\in\mathbb{N}}$ 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 $\|A\|_{\mathcal{L}(\mathcal{U},\mathcal{U})} = |\lambda_1|$.

If A is not self-adjoint, the decomposition in Theorem 2.2.4 does not hold any more. Instead, we can consider the so-called *singular value decomposition* of a compact linear operator.

Theorem 2.2.6. *Let $A \in \mathcal{K}(\mathcal{U}, \mathcal{V})$. Then there exists*

(a) a not-necessarily infinite null sequence $\{\sigma_j\}_{j\in\mathbb{N}}$ with $\sigma_1 \geqslant \sigma_2 \geqslant \ldots > 0$,

(b) an orthonormal basis $\{x_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$ of $\mathcal{N}(A)^\perp$,

(c) an orthonormal basis $\{y_j\}_{j\in\mathbb{N}} \subset \mathcal{V}$ of $\overline{\mathcal{R}(A)}$ with

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

Moreover, for all $u \in \mathcal{U}$ 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

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

Proof. Consider $B = A^*A$ and $C = AA^*$. Both B and C are compact, self-adjoint and even positive semidefinite, so that by Theorem 2.2.4 both admit a spectral representation and, by positive semidefiniteness, their eigenvalues are positive, i.e.

$$Bu = \sum_{j=1}^{\infty} \sigma_j^2 \langle u, x_j \rangle x_j \quad \forall u \in \mathcal{U}, \quad Cf = \sum_{j=1}^{\infty} \tilde{\sigma}_j^2 \langle f, y_j \rangle y_j \quad \forall f \in \mathcal{V},$$

where $\{x_j\}$ and $\{y_j\}$ are orthonormal bases of $\overline{\mathcal{R}(A^*A)}$ and $\overline{\mathcal{R}(AA^*)}$, respectively, and $\sigma_j, \tilde{\sigma}_j > 0$ for all j. As pointed out in Remark 2.0.2 and Lemma 2.0.3, we have $\overline{\mathcal{R}(A^*A)} = \overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp$ and, therefore, $\{x_j\}$ is also a basis of $\mathcal{N}(A)^\perp$. Analogously, $\{y_j\}$ is also a basis of $\overline{\mathcal{R}(A)}$.

Since $\tilde{\sigma}_j^2$ is an eigenvalue of C for the eigenvector y_j, we get that

$$\tilde{\sigma}_j^2 A^* y_j = A^*(\tilde{\sigma}_j^2 y_j) = A^* C y_j = A^* A A^* y_j = B A^* y_j$$

and therefore $\tilde{\sigma}_j^2$ is also an eigenvalue of B (for the eigenvector $A^* y_j$). Hence, with no loss of generality we can assume that $\tilde{\sigma}_j = \sigma_j$. We further observe that $\left\{\frac{A^* y_j}{\sigma_j}\right\}$ form an orthonormal basis of $\overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp$, since

$$\left\langle \frac{A^* y_j}{\sigma_j}, \frac{A^* y_k}{\sigma_k} \right\rangle = \frac{1}{\sigma_j \sigma_k} \langle y_j, AA^* y_k \rangle = \frac{1}{\sigma_j \sigma_k} \langle y_j, \sigma_k^2 y_k \rangle = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, we can choose $\{x_j\}$ to be

$$x_j = \sigma_j^{-1} A^* y_j$$

and we get that

$$A^* y_j = \sigma_j x_j.$$

We also observe that

$$A x_j = \sigma_j^{-1} A A^* y_j = \sigma_j^{-1} \sigma_j^2 y_j = \sigma_j y_j,$$

which proves (2.5).

Extending the basis $\{x_j\}$ of $\overline{\mathcal{R}(A^*)}$ to a basis of $\mathcal{U}$, we expand an arbitrary $u \in \mathcal{U}$ as $u = \sum_{j=1}^\infty \langle u, x_j \rangle x_j$ and, since $\mathcal{U} = \mathcal{N}(A) \oplus \overline{\mathcal{R}(A^*)}$ (Remark 2.0.2), obtain the singular value decompositions (2.6) – (2.7)

$$Au = \sum_{j=1}^\infty \sigma_j \langle u, x_j \rangle y_j \quad \forall u \in \mathcal{U}, \quad A^* f = \sum_{j=1}^\infty \sigma_j \langle f, y_j \rangle x_j \quad \forall f \in \mathcal{V}.$$

□

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

Theorem 2.2.7. *Let $A \in \mathcal{K}(\mathcal{U}, \mathcal{V})$ with singular system $\{(\sigma_j, x_j, y_j)\}_{j \in \mathbb{N}}$ and $f \in \mathcal{D}(A^\dagger)$. 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 $f \in \mathcal{D}(A^\dagger)$, $u^\dagger = A^\dagger f$ solves the normal equations

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

From Theorem 2.2.6 we know that

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

which implies that

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

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

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

□

The representation (2.8) makes it clear again that the Moore-Penrose inverse is unbounded. 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 2.2.8. *We say that the data f satisfy the Picard criterion, if*

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

Remark 2.2.9. 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.

It turns our that the Picard criterion also can be used to characterise elements in the range of the forward operator.

Theorem 2.2.10. *Let $A \in \mathcal{K}(\mathcal{U}, \mathcal{V})$ with singular system $\{(\sigma_j, x_j, y_j)\}_{j \in \mathbb{N}}$, and $f \in \overline{\mathcal{R}(A)}$. Then $f \in \mathcal{R}(A)$ if and only if the* Picard criterion

$$\sum_{j=1}^\infty \frac{\left|\langle f, y_j \rangle_\mathcal{V}\right|^2}{\sigma_j^2} < \infty \tag{2.11}$$

is met.

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

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

and therefore

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

Now let the Picard criterion (2.11) hold and define $u := \sum_{j=1}^\infty \sigma_j^{-1} \langle f, y_j \rangle_\mathcal{V} x_j \in \mathcal{U}$. It is well-defined by the Picard criterion (2.11) and we conclude

$$Au = \sum_{j=1}^\infty \sigma_j^{-1} \langle f, y_j \rangle_\mathcal{V} Ax_j = \sum_{j=1}^\infty \langle f, y_j \rangle_\mathcal{V} y_j = P_{\overline{\mathcal{R}(A)}} f = f\,,$$

which shows $f \in \mathcal{R}(A)$. □

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 2.2.11. *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 $\sigma_j \geqslant Cj^{-\gamma}$ 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 $\sigma_j \leqslant Cj^{-\gamma}$ for j sufficiently large.*

Example 2.2.12. Let us consider the example of differentiation again, as introduced in Section 1.1.1. The forward operator $A\colon L^2([0,1]) \to L^2([0,1])$ in this problem is given by

$$(Au)(t) = \int_0^t u(s)\,ds = \int_0^1 K(s,t)u(s)\,ds\,,$$

with $K\colon [0,1]\times[0,1] \to \mathbb{R}$ 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 1.1.2. Since the kernel K is square integrable, A is compact.

The adjoint operator A^* is given via

$$(A^*f)(s) = \int_0^1 K(t,s)f(t)\,dt = \int_s^1 v(t)\,dt\,. \tag{2.12}$$

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

$$\sigma^2 x(s) = (A^*Ax)(s) = \int_s^1 \int_0^t x(r)\,dr\,dt\,.$$

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

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

from which we conclude $x'(0) = 0$. Taking the derivative another time thus 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 $x'(0) = c_1 = 0$, we chose $c_1 = 0$ and σ such that $\cos(\sigma^{-1}) = 0$. Hence, we have

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

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

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

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

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

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

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

and formally differentiating the series, we obtain

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

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.

Chapter 3

Regularisation Theory

3.1 What is Regularisation?

We have seen that the Moore-Penrose inverse $A^\dagger$ is unbounded. 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 $f \in \mathcal{D}(A^\dagger)$ and all $f_\delta \in \mathcal{V}$ s.t. $\|f - f_\delta\|_{\mathcal{V}} \leqslant \delta$ as $\delta \to 0$.

Definition 3.1.1. *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ 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 $f \in \mathcal{D}(A^\dagger)$ as $\alpha \to 0$.

Definition 3.1.2. *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. As in the interesting cases the Moore–Penrose inverse may not be continuous we cannot expect that the norm of R_α stays bounded as $\alpha \to 0$. This is confirmed by the following results.

Theorem 3.1.3 (Banach–Steinhaus e.g. [8, p. 78], [26, p. 173])**.** *Let $\mathcal{U}, \mathcal{V}$ be Hilbert spaces and $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}(\mathcal{U}, \mathcal{V})$ a family of point-wise bounded operators, i.e. for all $u \in \mathcal{U}$ there exists a constant $C(u) > 0$ with $\sup_{j\in\mathbb{N}} \|A_j u\|_{\mathcal{V}} \leqslant C(u)$. Then*

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

Corollary 3.1.4 ([26, p. 174])**.** *Let $\mathcal{U}, \mathcal{V}$ be Hilbert spaces and $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then the following two conditions are equivalent:*

(a) There exists $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ such that

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

(b) There is a dense subset $\mathcal{X} \subset \mathcal{U}$ such that $\lim_{j\to\infty} A_j u$ exists for all $u \in \mathcal{X}$ and

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

Theorem 3.1.5. *Let $\mathcal{U}$, $\mathcal{V}$ be Hilbert spaces, $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and $\{R_\alpha\}_{\alpha>0}$ a linear regularisation as defined in Definition 3.1.2. If $A^\dagger$ is not continuous, $\{R_\alpha\}_{\alpha>0}$ cannot be uniformly bounded. In particular this implies the existence of an element $f \in \mathcal{V}$ with $\|R_\alpha f\| \to \infty$ for $\alpha \to 0$.*

Proof. We prove the theorem by contradiction and assume that $\{R_\alpha\}_{\alpha>0}$ is uniformly bounded. Hence, there exists a constant C with $\|R_\alpha\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} \leqslant C$ for all $\alpha > 0$. Due to Definition 3.1.1, we have $R_\alpha \to A^\dagger$ on $\mathcal{D}(A^\dagger)$. Since $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, by Corollary 3.1.4 we get that $A^\dagger \in \mathcal{L}(\mathcal{U}, \mathcal{V})$, which is a contradiction to the assumption that $A^\dagger$ is not continuous.

It remains to show the existence of an element $f \in \mathcal{V}$ with $\|R_\alpha f\|_\mathcal{V} \to \infty$ for $\alpha \to 0$. If such an element would not exist, we could conclude $\{R_\alpha\}_{\alpha>0} \subset \mathcal{L}(\mathcal{U}, \mathcal{V})$. However, Theorem 3.1.3 then implies that $\{R_\alpha\}_{\alpha>0}$ has to be uniformly bounded, which contradicts the first part of the proof. □

With the additional assumption that $\|AR_\alpha\|_{\mathcal{L}(\mathcal{U},\mathcal{V})}$ is bounded, we can even show that $R_\alpha f$ diverges for all $f \notin \mathcal{D}(A^\dagger)$.

Theorem 3.1.6. *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and $\{R_\alpha\}_{\alpha>0}$ be a linear regularisation of $A^\dagger$. If*

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

then $\|AR_\alpha f\|_\mathcal{U} \to \infty$ for $f \notin \mathcal{D}(A^\dagger)$.

Proof. Define $u_\alpha := R_\alpha f$ for $f \notin \mathcal{D}(A^\dagger)$. Assume that there exists a sequence $\alpha_k \to 0$ such that $\|u_{\alpha_k}\|_\mathcal{U}$ is uniformly bounded. Since bounded sets in a Hilbert space are weakly pre-compact, there exists a weakly convergent subsequence $u_{\alpha_{k_l}}$ with some limit $u \in \mathcal{U}$, cf. [15, Section 2.2, Theorem 2.1]. As continuous linear operators are also weakly continuous, we further have $Au_{\alpha_{k_l}} \rightharpoonup Au$. On the other hand, for any $f \in \mathcal{D}(A^\dagger)$ we have that $AR_\alpha f \to AA^\dagger f = P_{\overline{\mathcal{R}(A)}} f$. By Corollary 3.1.4 we then conclude that this also holds for any $f \in \mathcal{V}$, i.e. also for $f \notin \mathcal{D}(A^\dagger)$. Therefore, we get that $Au = P_{\overline{\mathcal{R}(A)}} f$. Since $\mathcal{V} = \overline{\mathcal{R}(A)} \oplus \mathcal{R}(A)^\perp$, we get that $f \in \mathcal{R}(A) \oplus \mathcal{R}(A)^\perp = \mathcal{D}(A^\dagger)$ in contradiction to the assumption $f \notin \mathcal{D}(A^\dagger)$. □

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 $f \in \mathcal{D}(A^\dagger)$ and all $f_\delta \in \mathcal{V}$ s.t. $\|f - f_\delta\|_\mathcal{V} \leqslant \delta$ 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

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

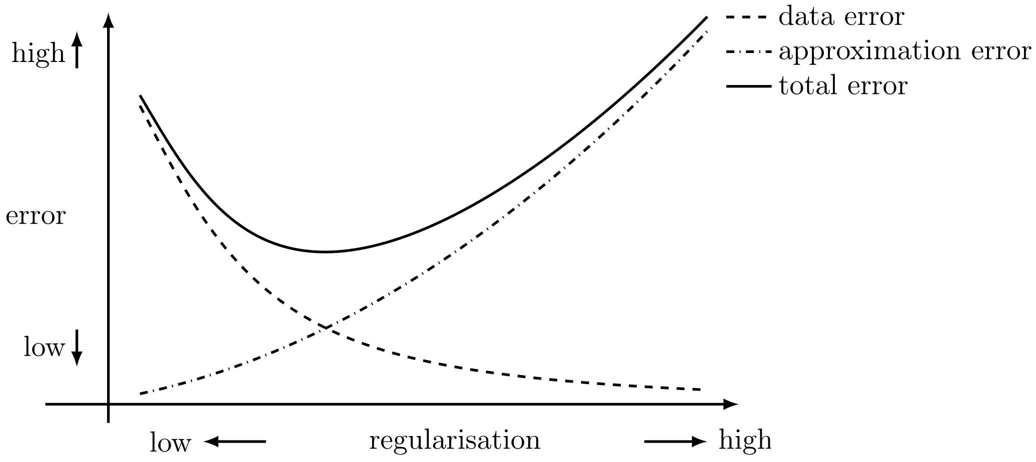

Figure 3.1: The *total error* between a regularised solution and the minimal norm solution decomposes into the *data error* and the *approximation error*. These two errors have opposing trends: For a small regularisation parameter α the error in the data gets amplified through the ill-posedness of the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

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 3.1.5. 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 3.2.1. *A function $\alpha\colon \mathbb{R}_{>0} \times \mathcal{V} \to \mathbb{R}_{>0}$, $(\delta, f_\delta) \mapsto \alpha(\delta, f_\delta)$ is called a* parameter choice rule. *We distinguish between*

(a) a priori parameter choice rules, which depend on δ only;

(b) a posteriori parameter choice rules, which depend on both δ and f_δ;

(c) 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 3.2.2. *Let $\{R_\alpha\}_{\alpha>0}$ be a regularisation of $A^\dagger$. If for all $f \in \mathcal{D}(A^\dagger)$ there exists a parameter choice rule $\alpha : \mathbb{R}_{>0} \times \mathcal{V} \to \mathbb{R}_{>0}$ such that*

$$\lim_{\delta\to 0} \sup_{f_\delta\colon \|f-f_\delta\|_{\mathcal{V}} \leqslant \delta} \|R_\alpha f_\delta - A^\dagger f\|_{\mathcal{U}} = 0 \tag{3.2}$$

and

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

then the pair (R_α, α) is called a convergent regularisation.

3.2.1 A priori parameter choice rules

First of all we want to discuss a priori parameter choice rules in more detail. Historically, they were the first to be studied. First let us show that for every regularisation an a priori parameter choice rule, and thus, a convergent regularisation, exists.

Theorem 3.2.3. *Let $\{R_\alpha\}_{\alpha>0}$ be a regularisation of $A^\dagger$, for $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then there exists an a priori parameter choice rule $\alpha = \alpha(\delta)$ such that (R_α, α) is a convergent regularisation.*

Proof. Let $f \in \mathcal{D}(A^\dagger)$ be arbitrary but fixed. Since $R_\alpha f \to A^\dagger f$, we can find a monotone increasing function $\gamma : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{\varepsilon\to 0} \gamma(\varepsilon) = 0$ such that for every $\varepsilon > 0$ we have

$$\left\| R_{\gamma(\varepsilon)} f - A^\dagger f \right\|_{\mathcal{U}} \leqslant \frac{\varepsilon}{2} .$$

As the operator $R_{\gamma(\varepsilon)}$ is continuous for fixed ε, there exists $\rho(\varepsilon) > 0$ with

$$\left\| R_{\gamma(\varepsilon)} g - R_{\gamma(\varepsilon)} f \right\|_{\mathcal{U}} \leqslant \frac{\varepsilon}{2} \text{ for all } g \in \mathcal{V} \text{ with } \|g - f\|_{\mathcal{V}} \leqslant \rho(\varepsilon) .$$

Without loss of generality we can assume ρ to be a continuous, strictly monotone increasing function with $\lim_{\varepsilon\to 0} \rho(\varepsilon) = 0$. Then, due to the inverse function theorem there exists a strictly monotone and continuous function ρ^{-1} on the range of ρ with $\lim_{\delta\to 0} \rho^{-1}(\delta) = 0$. We continuously extend ρ^{-1} on $\mathbb{R}_{>0}$ and define our a priori strategy as

$$\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{>0}, \quad \alpha(\delta) = \gamma(\rho^{-1}(\delta)) .$$

Then $\lim_{\delta\to 0} \alpha(\delta) = 0$ follows. Furthermore, for all $\varepsilon > 0$ there exists $\delta := \rho(\varepsilon)$, such that with $\alpha(\delta) = \gamma(\varepsilon)$

$$\left\| R_{\alpha(\delta)} f_\delta - A^\dagger f \right\|_{\mathcal{U}} \leqslant \left\| R_{\gamma(\varepsilon)} f_\delta - R_{\gamma(\varepsilon)} f \right\|_{\mathcal{U}} + \left\| R_{\gamma(\varepsilon)} f - A^\dagger f \right\|_{\mathcal{U}} \leqslant \varepsilon$$

follows for all $f_\delta \in \mathcal{V}$ with $\|f - f_\delta\|_{\mathcal{V}} \leqslant \delta$. Thus, (R_α, α) is a convergent regularisation method. □

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

Theorem 3.2.4. *Let $\{R_\alpha\}_{\alpha>0}$ be a linear regularisation, and $\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ an a priori parameter choice rule. Then (R_α, α) is a convergent regularisation method if and only if*

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

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

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

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

Hence, (R_α, α) is a convergent regularisation method.
$\Rightarrow$: Now let (R_α, α) 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_α, α) being a convergent regularisation method. If condition a) is violated, (3.3) is violated and hence, (R_α, α) is not a convergent regularisation method. If

condition a) is fulfilled but condition b) is violated, there exists a null sequence $\{\delta_k\}_{k\in\mathbb{N}}$ with $\delta_k\|R_{\alpha(\delta_k)}\|_{\mathcal{L}(\mathcal{V},\mathcal{U})} \geqslant C > 0$, and hence, we can find a sequence $\{g_k\}_{k\in\mathbb{N}} \subset \mathcal{V}$ with $\|g_k\|_{\mathcal{V}} = 1$ and $\delta_k\|R_{\alpha(\delta_k)}g_k\|_{\mathcal{U}} \geqslant \tilde{C}$ for some $\tilde{C}$. Let $f \in \mathcal{D}(A^\dagger)$ be arbitrary and define $f_k := f + \delta_k g_k$. Then we have on the one hand $\|f - f_k\|_{\mathcal{V}} \leqslant \delta_k$, 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 $f_\delta = f + g_k$ and thus, (R_α, α) 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 $\tilde{\alpha} = \alpha(C\delta)$ with any $C > 0$ also defines a convergent regularisation (for linear regularisations, it is a trivial corollary of Theorem 3.2.4). 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 $f \in \mathcal{D}(A^\dagger)$ and $f_\delta \in \mathcal{V}$ 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

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

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 $\mathcal{R}(A)$ is dense in $\mathcal{V}$, we get that $\mathcal{R}(A)^\perp = \{0\}$ due to Remark 2.0.2 and $\mu = 0$. Therefore, we ideally ensure that $\mathcal{R}(A)$ is dense.

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

Definition 3.2.5 (Morozov's discrepancy principle)**.** *Let $u_\alpha = R_\alpha f_\delta$ with $\alpha(\delta, f_\delta)$ chosen as follows*

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

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 regularization method [14, Chapter 4.3].

Practical a-posteriori regularisation strategies are usually designed as follows. We pick a null sequence $\{\alpha_j\}_{j\in\mathbb{N}}$ and iteratively compute $u_{\alpha_j} = R_{\alpha_j} f_\delta$ for $j \in \{1, \dots, j^*\}$, $j^* \in \mathbb{N}$, until $u_{\alpha_{j^*}}$ satisfies Morozov's discrepancy principle. This procedure is justified by the following theorem.

Theorem 3.2.6. *Let $\{R_\alpha\}_{\alpha>0}$ be a linear regularisation of $A^\dagger$ and $\{AR_\alpha\}_{\alpha>0}$ be uniformly bounded. Moreover, let $\mathcal{R}(A)$ be dense in $\mathcal{V}$, $f \in \mathcal{V}$ and let $\{\alpha_j\}_{j\in\mathbb{N}}$ be a null sequence. Then, for all $\delta > 0$ there exists a finite index $j^* \in \mathbb{N}$ such that the inequalities*

$$\|Au_{\alpha_{j^*}} - f_\delta\| \leqslant \eta\delta < \|Au_{\alpha_j} - f_\delta\|$$

are satisfied for all $j < j^*$.

Proof. We know that $\{AR_\alpha\}$ converges pointwise to $AA^\dagger = P_{\overline{\mathcal{R}(A)}}$ on $\mathcal{D}(A^\dagger)$, which together with the uniform boundedness assumption already implies pointwise convergence in $\mathcal{V}$, as we have already shown in the proof of Theorem 3.1.5. Hence, for all $f \in \mathcal{V}$ we can conclude that

$$\lim_{\alpha\to 0} \|Au_\alpha - f\| = \lim_{\alpha\to 0} \|AR_\alpha f - f\| = \|P_{\overline{\mathcal{R}(A)}} f - f\| = 0.$$

□

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 yield convergent regularisations only for well-posed problems, as the following result, known as the Bakushinskii veto [5], demonstrates.

Theorem 3.2.7. *Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and $\{R_\alpha\}$ be a regularization for $A^\dagger$. Let $\alpha = \alpha(f_\delta)$ be a parameter choice rule such that (R_α, α) is a convergent regularization. Then $A^\dagger$ is continuous from $\mathcal{V}$ to $\mathcal{U}$.*

Proof. Since, for an arbitrary $f \in \mathcal{V}$, $R_{\alpha(f)} f$ only depends on f (and not on any additional parameters), we can as well define an operator $R' \colon \mathcal{V} \to \mathcal{U}$ such that $R'f := R_{\alpha(f)} f$ for all $f \in \mathcal{V}$. By the definition of a convergent regularisation (Def. 3.2.2) we get that for any $f \in \mathcal{D}(A^\dagger)$

$$\lim_{\delta\to 0} \sup_{f_\delta \colon \|f-f_\delta\|_\mathcal{V} \leqslant \delta} \|R_{\alpha(f_\delta)} f_\delta - A^\dagger f\|_\mathcal{U} = \lim_{\delta\to 0} \sup_{f_\delta \colon \|f-f_\delta\|_\mathcal{V} \leqslant \delta} \|R' f_\delta - A^\dagger f\|_\mathcal{U} = 0.$$

Taking $f_\delta = f$ for all δ, we get that $R'f = A^\dagger f$ for all $f \in \mathcal{D}(A^\dagger)$. Therefore, for any $f \in \mathcal{D}(A^\dagger)$ and for every sequence $\{f_j\}_{j\in\mathbb{N}} \subset \mathcal{V}$ converging to f we get that

$$\lim_{j\to\infty} \|R_{\alpha(f)} f_j - A^\dagger f\| = \lim_{j\to\infty} \|R' f_j - R' f\| = 0,$$

i.e. R' is continuous on $\mathcal{D}(A^\dagger)$ in $\mathcal{V}$. In particular, we get that on $\mathcal{D}(A^\dagger)$, $A^\dagger = R'$ is continuous. Since $A^\dagger$ is continuous, by Theorem 2.1.11 we get that $\mathcal{R}(A)$ is closed and $\mathcal{D}(A^\dagger) = \mathcal{V}$, hence $A^\dagger$ is continuous from $\mathcal{V}$ to $\mathcal{U}$. □

Remark 3.2.8. Since $\mathcal{D}(A^\dagger) = \mathcal{V}$, we get that $R' = A^\dagger$ on the whole space $\mathcal{V}$ and we conclude that $R'(\cdot) = R_{\alpha(\cdot)}(\cdot)$ has to be linear. However, we did not assume linearity of R_α or anything at all about the parameter choice rule $\alpha(f_\delta)$ and cannot expect linearity of R', which is another contradiction.

3.3 Spectral Regularisation

Recall the spectral representation (2.8) of the Moore-Penrose inverse $A^\dagger$

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

where $\{(\sigma_j, x_j, y_j)\}$ is the singular system of A.

The source of ill-posedness of $A^\dagger$ are the eigenvalues $1/\sigma_j$, which explode as $j \to \infty$, since $\sigma_j \to 0$ as $j \to \infty$. Let us construct a regularisation by modifying these eigenvalues as follows

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

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

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

Theorem 3.3.1. *Let $g_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ 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 $f \in \mathcal{D}(A^\dagger)$.

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

$$R_\alpha f - A^\dagger f = \sum_{j=1}^{\infty} \left(g_\alpha(\sigma_j) - \frac{1}{\sigma_j} \right) \langle f, y_j \rangle_{\mathcal{V}} \, x_j = \sum_{j=1}^{\infty} \left(\sigma_j g_\alpha(\sigma_j) - 1 \right) \langle u^\dagger, x_j \rangle_{\mathcal{U}} \, x_j \,.$$

Consider

$$\| R_\alpha f - A^\dagger f \|_{\mathcal{U}}^2 = \sum_{j=1}^{\infty} (\sigma_j g_\alpha(\sigma_j) - 1)^2 \left| \langle u^\dagger, x_j \rangle_{\mathcal{U}} \right|^2 .$$

From (3.7) we can conclude

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

whilst

$$\sum_{j=1}^{\infty} (1 + \gamma^2) \left| \langle u^\dagger, x_j \rangle_{\mathcal{U}} \right|^2 = (1 + \gamma^2) \| u^\dagger \|^2 < +\infty.$$

Therefore, by the reverse Fatou lemma we get the following estimate

$$\begin{aligned}\limsup_{\alpha\to 0}\left\|R_\alpha f - A^\dagger f\right\|_{\mathcal{U}}^2 &= \limsup_{\alpha\to 0}\sum_{j=1}^\infty (\sigma_j g_\alpha(\sigma_j)-1)^2\left(\langle u^\dagger, x_j\rangle_{\mathcal{U}}\right)^2\\ &\leqslant \sum_{j=1}^\infty\left(\limsup_{\alpha\to 0}\sigma_j g_\alpha(\sigma_j)-1\right)^2\left|\langle u^\dagger, x_j\rangle_{\mathcal{U}}\right|^2 = 0\,,\end{aligned}$$

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

Theorem 3.3.2. *Let the assumptions of Theorem 3.3.1 hold and let $\alpha = \alpha(\delta)$ be an a-priori parameter choice rule. Then $(R_{\alpha(\delta)}, \alpha(\delta))$ with R_α as defined in* (3.5) *is a convergent regularisation method if*

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

Proof. The result follows immediately from $\|R_{\alpha(\delta)}\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} \leqslant C_{\alpha(\delta)}$ and Theorem 3.2.4. □

3.3.1 Truncated singular value decomposition

As a first example for a spectral regularisation of the form (3.5) we want to 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

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

for all $f \in \mathcal{V}$. 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 vectors of compact operators.

Let $A \in \mathcal{K}(\mathcal{U}, \mathcal{V})$ with singular system $\{\sigma_j, x_j, y_j\}_{j\in\mathbb{N}}$, and choose for $\delta > 0$ an index function $j^* : \mathbb{R}_+ \to \mathbb{N}$ 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

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

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 all singular vectors 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 eigenvectors 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

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

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 3.3.3. 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

$$\min_{u \in \mathcal{U}} \|Au - f\|^2 + \alpha \|u\|^2. \tag{3.15}$$

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.

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

Chapter 4

Variational Regularisation

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

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

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 *regularistaion 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 notations $\mathcal{F}(Au, f)$ for the fidelity function and $\mathcal{J}(u)$ for the regulariser, we can formally write down the variational regularisation problem as follows

$$\min_{u \in \mathcal{U}} \mathcal{F}(Au, f) + \alpha \mathcal{J}(u), \tag{4.1}$$

i.e.

$$R_\alpha f \in \arg\min_{u \in \mathcal{U}} \mathcal{F}(Au, f) + \alpha \mathcal{J}(u).$$

In this chapter, we will study the properties of (4.1) for different choices of $\mathcal{F}$ and $\mathcal{J}$, 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 (as Hilbert spaces) but they may not have an inner product. For every Banach space $\mathcal{U}$, we can define the space of linear and continuous functionals which is called the *dual space* $\mathcal{U}^*$ of $\mathcal{U}$, i.e. $\mathcal{U}^* := \mathcal{L}(\mathcal{U}, \mathbb{R})$. Let $u \in \mathcal{U}$ and $p \in \mathcal{U}^*$, then we usually write the *dual product* $\langle p, u \rangle$ instead of $p(u)$. Moreover, for any $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ there exists a unique operator $A^* \colon \mathcal{V}^* \to \mathcal{U}^*$, called the *adjoint* of A such that for all $u \in \mathcal{U}$ and $p \in \mathcal{V}^*$ 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. $A^* p \in \mathcal{U}^*$ and $u \in \mathcal{U}$.

The dual space of a Banach space $\mathcal{U}$ can be equipped with the following norm

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

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 $\mathcal{U}$ and denote it with $\mathcal{U}^{**} := (\mathcal{U}^*)^*$. As every $u \in \mathcal{U}$ defines a continuous and linear mapping on the dual space $\mathcal{U}^*$ by

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

the mapping $E \colon \mathcal{U} \to \mathcal{U}^{**}$ 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 $\mathcal{U}$ *reflexive*. Examples of reflexive Banach spaces include Hilbert spaces and L^q, ℓ^q spaces with $1 < q < \infty$. We call the space $\mathcal{U}$ *separable* if there exists a set $\mathcal{X} \subset \mathcal{U}$ of at most countable cardinality such that $\overline{\mathcal{X}} = \mathcal{U}$.

A problem in infinite dimensional spaces is that bounded sequences may fail to have convergent subsequences. An example is for instance in ℓ^2 the sequence $\{u^k\}_{k\in\mathbb{N}} \subset \ell^2, u^k_j = 1$ 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 $\mathcal{U}$. We say that $\{u^k\}_{k\in\mathbb{N}} \subset \mathcal{U}$ *converges weakly* to $u \in \mathcal{U}$ if and only if for all $p \in \mathcal{U}^*$ the sequence of real numbers $\{\langle p, u^k \rangle\}_{k\in\mathbb{N}}$ 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 $\mathcal{U}^*$ 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 $\{p^k\}_{k\in\mathbb{N}} \subset \mathcal{U}^*$ *converges in weak-$*$* to $p \in \mathcal{U}^*$ if and only if

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

and we denote weak-$*$ convergence by $p^k \overset{*}{\rightharpoonup} p$. Similarly, for any topology τ on $\mathcal{U}$ we denote the convergence in that topology by $u^k \overset{\tau}{\to} u$.

With these two new notions of convergence, we can solve the problem of bounded sequences:

Theorem 4.1.1 (Sequential Banach-Alaoglu Theorem, e.g. [22, p. 70] or [24, p. 141])**.** *Let $\mathcal{U}$ be a separable normed vector space. Then every bounded sequence $\{u^k\}_{k\in\mathbb{N}} \subset \mathcal{U}^*$ has a weak-* convergent subsequence.*

Theorem 4.1.2 ([26, p. 64])**.** *Each bounded sequence $\{u^k\}_{k\in\mathbb{N}}$ in a reflexive Banach space $\mathcal{U}$ 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 4.1.3. *Let $\mathcal{U}$ be a Banach space with topology $\tau_{\mathcal{U}}$. The functional $E \colon \mathcal{U} \to \bar{\mathbb{R}}$ is said to be* sequentially lower semi-continuous with respect to $\tau_{\mathcal{U}}$ *($\tau_{\mathcal{U}}$-l.s.c.) at $u \in \mathcal{U}$ if*

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

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

Remark 4.1.4. For topologies that are not induced by a metric we have to differ between a topological property and its sequential version, e.g. continuous and sequentially continuous. If the topology is induced by a metric, then these two are the same. However, for instance the weak and weak-$*$ topology are generally not induced by a metric.

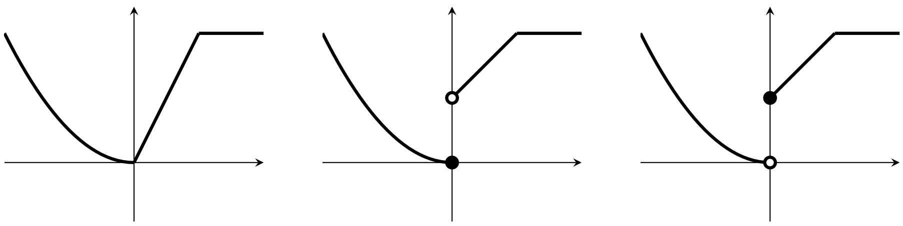

Figure 4.1: Visualisation of lower semi-continuity. The solid dot at a jump indicates the value that the function takes. The function on the left is continuous and thus lower semi-continuous. The functions in the middle and on the right are discontinuous. While the function in the middle is lower semi-continuous,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

Example 4.1.5. The functional $\|\cdot\|_1 : \ell^2 \to \bar{\mathbb{R}}$ 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 $\{u^j\}_{j\in\mathbb{N}} \subset \ell^2$ be a weakly convergent sequence with $u^j \rightharpoonup u \in \ell^2$. We have with $\delta_k : \ell^2 \to \mathbb{R}, \langle \delta_k, v\rangle = v_k$ that for all $k \in \mathbb{N}$

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

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 it is not clear whether both the left and the right hand side are finite. □

4.1.2 Convex analysis

Infinity calculus

We will look at functionals $E : \mathcal{U} \to \bar{\mathbb{R}}$ whose range is modelled to be the *extended real line* $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ 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 $x \in \mathbb{R}$ (similarly, $x > -\infty$ for all $x \in \mathbb{R}$). This is useful to model constraints: for instance, if we were trying to minimise $E : [-1, \infty) \to \mathbb{R}, x \mapsto x^2$ we could remodel this minimisation problem by $\tilde{E} : \mathbb{R} \to \bar{\mathbb{R}}$

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

Obviously both functionals have the same minimiser but $\tilde{E}$ is defined on a vector space and not only on a subset. This has two important consequences: on the on 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 $\tilde{E}$ 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 $\bar{\mathbb{R}}$ as we are used to on the real line $\mathbb{R}$ but the operations with $\pm\infty$ need yet to be defined.

Definition 4.1.6. *The extended real line is defined as* $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ *with the following rules that hold for any* $x \in \mathbb{R}$ *and* $\lambda > 0$*:*

$$x + \infty := \infty + x := \infty \qquad \lambda \cdot \infty := \infty \cdot \lambda := \infty$$
$$x/\infty := 0 \qquad \infty + \infty := \infty\,.$$

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

$$\infty - \infty \text{ and } \quad \infty \cdot \infty\,.$$

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

Definition 4.1.7 (Characteristic function)**.** *Let* $\mathcal{C} \subset \mathcal{U}$ *be a set. The function* $\chi_{\mathcal{C}} \colon \mathcal{U} \to \bar{\mathbb{R}}$*,*

$$\chi_{\mathcal{C}}(u) = \begin{cases} 0 & u \in \mathcal{C} \\ \infty & u \in \mathcal{U} \setminus \mathcal{C} \end{cases}$$

is called the characteristic function of the set $\mathcal{C}$*.*

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

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

Definition 4.1.8. *Let* $\mathcal{U}$ *be a vector space and* $E \colon \mathcal{U} \to \bar{\mathbb{R}}$ *a functional. Then the* effective domain *of* E *is*

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

Definition 4.1.9. *A functional* E *is called* proper *if the effective domain* $\operatorname{dom}(E)$ *is not empty.*

Convexity

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

Definition 4.1.10. *Let* $\mathcal{U}$ *be a vector space. A subset* $\mathcal{C} \subset \mathcal{U}$ *is called* convex*, if* $\lambda u + (1 - \lambda)v \in \mathcal{C}$ *for all* $\lambda \in (0, 1)$ *and all* $u, v \in \mathcal{C}$*.*

Definition 4.1.11. *A functional* $E \colon \mathcal{U} \to \bar{\mathbb{R}}$ *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.*

Obviously, strict convexity implies convexity.

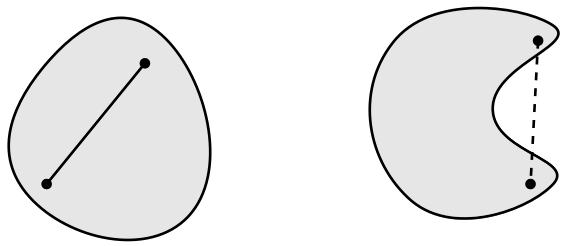

Figure 4.2: Example of a convex set (left) and non-convex set (right).

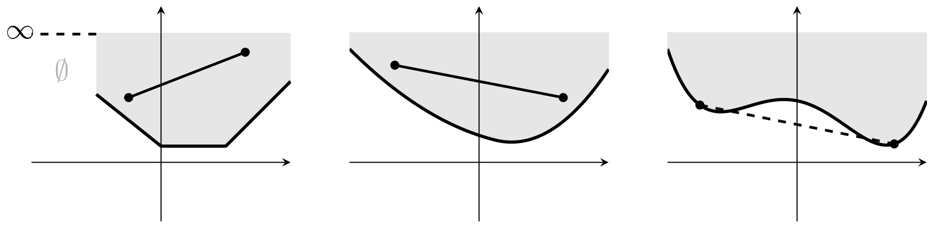

Figure 4.3: Example of a convex function (left), a strictly convex function (middle) and a non-convex function (right).

Example 4.1.12. The absolute value function $\mathbb{R} \to \mathbb{R}, x \mapsto |x|$ is convex but not strictly convex while the quadratic function $x \mapsto x^2$ is strictly convex. For other examples, see Figure 4.3.

Example 4.1.13. The characteristic function $\chi_{\mathcal{C}}(u)$ is convex if and only if $\mathcal{C}$ is a convex set. To see the convexity, let $u, v \in \mathrm{dom}(\chi_{\mathcal{C}}) = \mathcal{C}$. Then by the convexity of $\mathcal{C}$ the convex combination $\lambda u + (1-\lambda)v$ is as well in $\mathcal{C}$ and both the left and the right hand side of the desired inequality are zero.

Lemma 4.1.14. *Let $\alpha \geqslant 0$ and $E, F \colon \mathcal{U} \to \bar{\mathbb{R}}$ be two convex functionals. Then $E + \alpha F \colon \mathcal{U} \to \bar{\mathbb{R}}$ 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 4.1.15. *Let $E \colon \mathcal{U} \to \bar{\mathbb{R}}$ be a functional. The functional $E^* \colon \mathcal{U}^* \to \bar{\mathbb{R}}$,*

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

is called the Fenchel conjugate of E.

Theorem 4.1.16 ([13, Prop. 4.1])**.** *For any functional $E \colon \mathcal{U} \to \bar{\mathbb{R}}$ the following inequality holds:*

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

If E is proper, lower-semicontinuous (see Def. 4.1.3) and convex, then

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

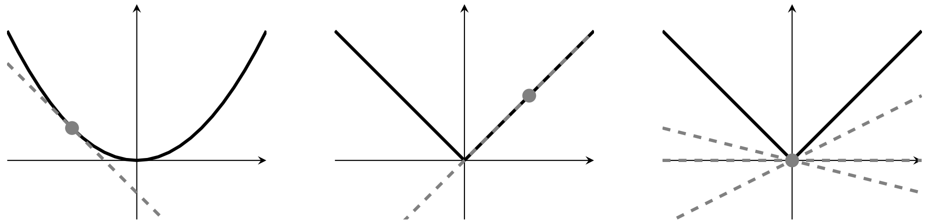

Figure 4.4: Visualisation of the subdifferential. Linear approximations of the functional have to lie completely underneath the function. For points where the function is not differentiable there may be more than one such approximation.

Subgradients

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

Definition 4.1.17. *A functional $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ is called* subdifferentiable *at $u \in \mathcal{U}$, if there exists an element $p \in \mathcal{U}^*$ such that*

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

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

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

is called subdifferential *of E at u.*

Remark 4.1.18. Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ be a convex functional. Then the subdifferential is non-empty at all $u \in \mathrm{dom}(E)$. If $\mathrm{dom}(E) \neq \emptyset$, then for all $u \notin \mathrm{dom}(E)$ the subdifferential is empty, i.e. $\partial E(u) = \emptyset$.

Theorem 4.1.19 ([3, Thm. 7.13])**.** *Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ be a proper convex function and $u \in \mathrm{dom}(E)$. Then $\partial E(u)$ is a weak-$*$ compact convex subset of $\mathcal{U}^*$.*

For differentiable functions the subdifferential consists of just one element – the derivative. For non-differentiable functionals the subdifferential is multivalued; we want to consider the subdifferential of the absolute value function as an illustrative example.

Example 4.1.20. Let $E\colon \mathbb{R} \to \mathbb{R}$ 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 4.4.

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

Theorem 4.1.21 ([13, Prop. 5.6])**.** *Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ and $F\colon \mathcal{U} \to \bar{\mathbb{R}}$ be proper l.s.c. convex functions and suppose $\exists u \in \mathrm{dom}(E) \cup \mathrm{dom}(F)$ 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 4.1.22. *An element $u \in \mathcal{U}$ is a minimiser of the functional $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ if and only if $0 \in \partial E(u)$.*

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

Bregman distances

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

Definition 4.1.23. *Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ be a convex functional. Moreover, let $u, v \in \mathcal{U}, E(v) < \infty$ and $q \in \partial E(v)$. Then the* (generalised) Bregman distance *of E between u and v is defined as*

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

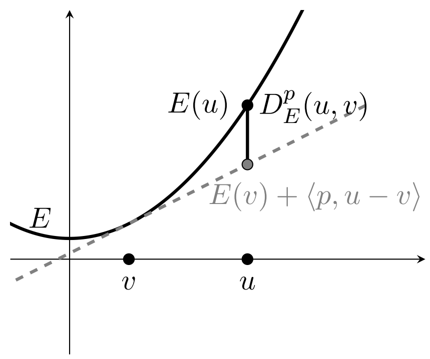

Figure 4.5: Visualization of the Bregman distance.

Remark 4.1.24. It is easy to check that a Bregman distance somewhat resembles a metric as for all $u, v \in \mathcal{U}, q \in \partial E(v)$ we have that $D_E^q(u, v) \geqslant 0$ and $D_E^q(v, v) = 0$. There are functionals where the Bregman distance (up to a square root) is actually a metric; e.g. $E(u) := \frac{1}{2}\|u\|_{\mathcal{U}}^2$ for Hilbert space $\mathcal{U}$, then $D_E^q(u, v) = \frac{1}{2}\|u - v\|_{\mathcal{U}}^2$. However, in general, Bregman distances are not symmetric and $D_E^q(u, v) = 0$ does not imply $u = v$, as you will see on the example sheets.

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

Definition 4.1.25. *Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ be a convex functional. Moreover, let $u, v \in \mathcal{U}, E(u) < \infty, E(v) < \infty$, $q \in \partial E(v)$ and $p \in \partial E(u)$. Then the* symmetric Bregman distance *of E between u and v is defined as*

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

Absolutely one-homogeneous functionals

Definition 4.1.26. *A functional $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ is called absolutely one-homogeneous if*

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

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

$$E(u) = (p, u).$$

Proof. Left as exercise. □

Remark 4.1.28. The Bregman distance $D_E^p(v, u)$ in this case can be written as follows:

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

Proposition 4.1.29. *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 4.1.30. *For any $u \in \mathcal{U}$, $p \in \partial E(u)$ if and only if $p \in \partial E(0)$ and $E(u) = (p, u)$.*

4.1.3 Minimisers

Definition 4.1.31. *Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ be a functional. We say that $u^* \in \mathcal{U}$ solves the minimisation problem*

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

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

Definition 4.1.32. *A functional $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ is called* bounded from below *if there exists a constant $C > -\infty$ such that for all $u \in \mathcal{U}$ we have $E(u) \geqslant C$.*

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

Existence

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 4.1.33. *A functional $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ is called* coercive*, if for all $\{u_j\}_{j \in \mathbb{N}}$ with $\|u_j\|_{\mathcal{U}} \to \infty$ we have $E(u_j) \to \infty$.*

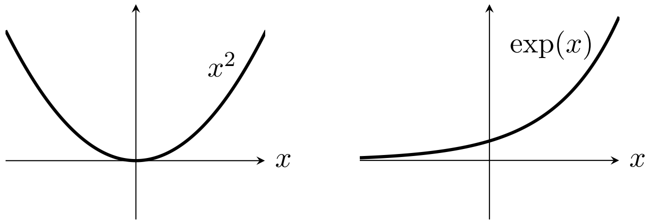

Figure 4.6: While the coercive function on the left has a minimiser, it is easy to see that the non-coercive function on the right does not have a minimiser.

Remark 4.1.34. Coercivity is equivalent to its negated statement which is "if the function values $\{E(u_j)\}_{j\in\mathbb{N}} \subset \mathbb{R}$ are bounded, so is the sequence $\{u_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$".

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

Lemma 4.1.35. *Let $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ be a proper, coercive functional and bounded from below. Then the infimum $\inf_{u\in\mathcal{U}} E(u)$ exists in $\mathbb{R}$, there are minimising sequences, i.e. $\{u_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$ with $E(u_j) \to \inf_{u\in\mathcal{U}} E(u)$, 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 $\{u_j\}_{j\in\mathbb{N}}$ be any minimising sequence, i.e. $E(u_j) \to \inf_u E(u)$. Then there exists a $j_0 \in \mathbb{N}$ 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 4.1.34. Including a finite number of elements does not change its boundedness which proves the assertion. $\square$

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

Proof. From Lemma 4.1.35 we know that $\inf_{u\in\mathcal{U}} E(u)$ is finite, minimising sequences exist and that they are bounded. Let $\{u_j\}_{j\in\mathbb{N}} \in \mathcal{U}$ be a minimising sequence. Thus, from the assumption on the topology $\tau_\mathcal{U}$ there exists a subsequence $\{u_{j_k}\}_{k\in\mathbb{N}}$ and $u^* \in \mathcal{U}$ with $u_{j_k} \stackrel{\tau_\mathcal{U}}{\to} u^*$ for $k \to \infty$. From the sequential lower semi-continuity of E we obtain

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

which shows that $E(u^*) < \infty$ and $E(u^*) \leqslant E(u)$ for all $u \in \mathcal{U}$; thus u^* minimises E. $\square$

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

Corollary 4.1.37. *Let $\mathcal{U}$ be a reflexive Banach space and $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ 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 4.1.36, as in reflexive Banach spaces bounded sequences have weakly convergent subsequences, see Theorem 4.1.2. □

Remark 4.1.38. For convex functionals on reflexive Banach spaces, 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. [13, Corollary 2.2., p. 11] or [6, p. 149] for Hilbert spaces).

Remark 4.1.39. 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.

Uniqueness

Theorem 4.1.40. *Assume that the functional $E\colon \mathcal{U} \to \bar{\mathbb{R}}$ 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 4.1.41. 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.

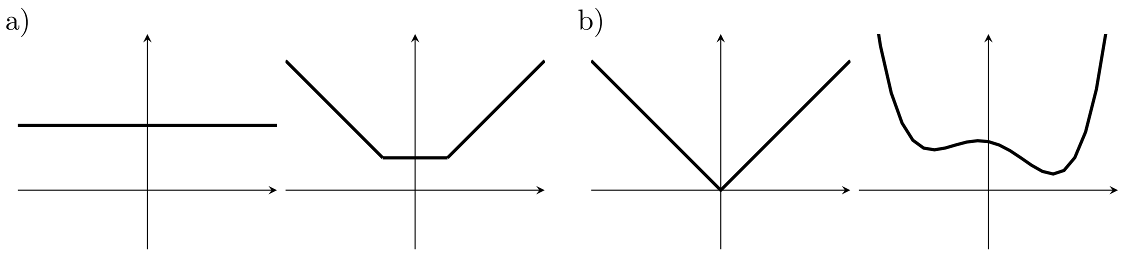

Figure 4.7: a) Convex functions may not have a unique minimiser. b) Neither strict convexity nor convexity is necessary for the uniqueness of a minimiser.

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 $A\colon \mathcal{U} \to \mathcal{V}$ is a linear bounded operator and $\mathcal{U}$ and $\mathcal{V}$ are Banach spaces (and not Hilbert spaces as in Chapter 3). 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 data term and regularisation functional. To this end, we need to extend the definitions of a least-squares solution and a minimal-norm solution (Def. 2.1.1) to an arbitrary data term and regularisation term.

Definition 4.2.1 ($\mathcal{J}$-minimising solutions)**.** *Suppose that the fidelity term is such that the optimisation problem*

$$\min_{u \in \mathcal{U}} \mathcal{F}(Au, f) \tag{4.5}$$

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

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

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

Remark 4.2.2. In order to simplify the presentation, we will assume that equation (4.4) has a solution with a finite value of $\mathcal{J}$, i.e. there exists at least one element $u^\dagger$ such that $Au^\dagger = f$ and $\mathcal{J}(u^\dagger) < +\infty$. With the natural assumption that $\mathcal{F}(f, g) \geqslant 0$ for all $f, g \in \mathcal{V}$ and $\mathcal{F}(f, f) = 0$ we get that problem (4.5) is solvable and its optimal value is zero.

Remark 4.2.3. Even if problem (4.5) is solvable, a $\mathcal{J}$-minimising solution may not exist. 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 choses 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 [7]). We will not explicitly mention this, stating all results for just *a* $\mathcal{J}$-minimising solution.

The next theorem states the main result of this chapter.

Theorem 4.2.4. *Let $\mathcal{U}$ and $\mathcal{V}$ be Banach spaces and $\tau_{\mathcal{U}}$ and $\tau_{\mathcal{V}}$ some topologies (not necessarily induced by the norm) in $\mathcal{U}$ and $\mathcal{V}$, respectively. Suppose that problem* (4.4) *is solvable and the solution has a finite value of $\mathcal{J}$. Assume also that*

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

(ii) $\mathcal{J}\colon \mathcal{U} \to \bar{\mathbb{R}}_+$ *is proper, $\tau_{\mathcal{U}}$-l.s.c. and its non-empty sublevel-sets* $\{u \in \mathcal{U}\colon \mathcal{J}(u) \leqslant C\}$ *are $\tau_{\mathcal{U}}$-sequentially compact;*

(iii) $\mathcal{F}\colon \mathcal{V} \times \mathcal{V} \to \bar{\mathbb{R}}_+$ *is proper, $\tau_{\mathcal{V}}$-l.s.c. in the first argument and norm-l.s.c. in the second one and satisfies*

$$\mathcal{F}(f, f) = 0 \quad \text{and} \quad \mathcal{F}(f, f_\delta) \leqslant C(\delta) \to 0 \quad \text{as } \delta \to 0$$

for all $f \in \mathcal{V}$ and all $f_\delta \in \mathcal{V}$ such that $\|f_\delta - f\| \leqslant \delta$;

(iv) there exists an a priori parameter choice rule $\alpha = \alpha(\delta)$ *such that* $\lim_{\delta\to 0}\alpha(\delta) = 0$ *and* $\lim_{\delta\to 0} C(\delta)/\alpha(\delta) = 0$.

Then

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

(ii') for any fixed $\alpha > 0$ *and* $f_\delta \in \mathcal{V}$ *there exists a minimiser* $u^\alpha_\delta \in \arg\min_{u\in\mathcal{U}} \mathcal{F}(Au, f_\delta) + \alpha\mathcal{J}(u)$;

(iii') the parameter choice rule $\alpha = \alpha(\delta)$ *from Assumption(iv) guarantees that* $u_\delta := u_\delta^{\alpha(\delta)} \overset{\tau_\mathcal{U}}{\to} u^\dagger_{\mathcal{J}}$ *as* $\delta \to 0$ *(possibly, along a subsequence) and* $\mathcal{J}(u_\delta) \to \mathcal{J}(u^\dagger_{\mathcal{J}})$.

Proof. (i') Under the assumptions made, optimisation problem (4.5) has at least one solution, i.e. the set of minimisers is non-empty. Denote it by $\mathcal{M}_\mathcal{F}(f)$. Let us show that $\mathcal{M}_\mathcal{F}(f)$ is $\tau_\mathcal{U}$-closed. Consider a sequence $\{u_n\} \subset \mathcal{M}_\mathcal{F}(f)$ such that $u_n \overset{\tau_\mathcal{U}}{\to} \bar{u}$. Since A is $\tau_\mathcal{U} \to \tau_\mathcal{V}$ continuous, we get that $Au_n \overset{\tau_\mathcal{V}}{\to} A\bar{u}$. Since $\mathcal{F}(\cdot,\cdot)$ is $\tau_\mathcal{V}$-l.s.c. in the first argument, we get that

$$\mathcal{F}(A\bar{u}, f) \leqslant \liminf_{n\to\infty} \mathcal{F}(Au_n, f) = 0,$$

since all u_n are minimisers of (4.5). Since $\mathcal{F}(\cdot,\cdot) \geqslant 0$, we get that $\mathcal{F}(A\bar{u}, f) = 0$ and $\bar{u}$ is a minimiser of (4.5), hence $\mathcal{M}_\mathcal{F}(f)$ is $\tau_\mathcal{U}$-closed.

A $\mathcal{J}$-minimsing solution solves the following problem

$$\min_{u\in\mathcal{M}_\mathcal{F}(f)} \mathcal{J}(u).$$

Since $\mathcal{J}$ is bounded from below, the infimum in this problem exists and we denote it by $\mathcal{J}_{min}$. Consider any minimising sequence $\{u_k\}$. By Assumption (ii), the sublevel sets of $\mathcal{J}$ are $\tau_\mathcal{U}$-sequentially compact and u_k contains a $\tau_\mathcal{U}$-converging subsequence $u_{k_j} \overset{\tau_\mathcal{U}}{\to} \tilde{u}$ as $j \to \infty$. Since $\mathcal{M}_\mathcal{F}(f)$ is $\tau_\mathcal{U}$-closed, $\tilde{u} \in \mathcal{M}_\mathcal{F}(f)$. Since $\mathcal{J}$ is $\tau_\mathcal{U}$-l.s.c., we get that

$$\mathcal{J}(\tilde{u}) \leqslant \liminf_{j\to\infty} \mathcal{J}(u_{k_j}) = \mathcal{J}_{min}.$$

Therefore, $\mathcal{J}(\tilde{u}) = \mathcal{J}_{min}$ and $\tilde{u}$ is a $\mathcal{J}$-minimising solution, which we from now on denote by $u^\dagger_{\mathcal{J}}$.

(ii') For a fixed $\alpha > 0$ and $f_\delta \in \mathcal{V}$ consider the following optimisation problem

$$\min_{u\in\mathcal{U}} \mathcal{F}(Au, f_\delta) + \alpha\mathcal{J}(u). \tag{4.6}$$

Comparing the value of the objective function for any minimising sequence $\{u_n\}$ and the $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$, we get that

$$\mathcal{J}(u_n) \leqslant \frac{1}{\alpha}\mathcal{F}(Au^\dagger_{\mathcal{J}}, f_\delta) + \mathcal{J}(u^\dagger_{\mathcal{J}}) = \text{const}.$$

By the sequential $\tau_\mathcal{U}$-compactness of the sublevel sets of $\mathcal{J}$ we get that $\{u_n\}$ contains a $\tau_\mathcal{U}$-converging subsequence $u_{n_j} \overset{\tau_\mathcal{U}}{\to} \hat{u}$. By Assumption (i) we get that $Au_{n_j} \overset{\tau_\mathcal{V}}{\to} A\hat{u}$. Since $\mathcal{F}(\cdot,\cdot)$ is $\tau_\mathcal{V}$-l.s.c in the first argument and $\mathcal{J}(\cdot)$ is $\tau_\mathcal{U}$-l.s.c., we get that

$$\mathcal{F}(A\hat{u}, f_\delta) + \alpha\mathcal{J}(\hat{u}) \leqslant \liminf_{j\to\infty} \mathcal{F}(Au_n, f_\delta) + \alpha\mathcal{J}(u_n) = \inf_{u\in\mathcal{U}} \mathcal{F}(Au, f_\delta) + \alpha\mathcal{J}(u).$$

Therefore, $\hat{u}$ is a minimiser in (4.6), denoted further by u^α_δ.

(iii') Let us study the behaviour of u_δ^α when $\delta \to 0$ and α is chosen according to the a priori parameter choice rule $\alpha = \alpha(\delta)$ from Assumption (iv). Denote $u_\delta^{\alpha(\delta)}$ by u_δ. Since u_δ solves (4.6) with $\alpha = \alpha(\delta)$, we get that

$$\mathcal{F}(Au_\delta, f_\delta) + \alpha(\delta)\mathcal{J}(u_\delta) \leqslant \mathcal{F}(Au_{\mathcal{J}}^\dagger, f_\delta) + \alpha(\delta)\mathcal{J}(u_{\mathcal{J}}^\dagger) \tag{4.7}$$

and, since $\mathcal{F}(\cdot,\cdot) \geqslant 0$,

$$\alpha(\delta)\mathcal{J}(u_\delta) \leqslant \mathcal{F}(Au_{\mathcal{J}}^\dagger, f_\delta) + \alpha(\delta)\mathcal{J}(u_{\mathcal{J}}^\dagger) = \mathcal{F}(f, f_\delta) + \alpha(\delta)\mathcal{J}(u_{\mathcal{J}}^\dagger) \leqslant C(\delta) + \alpha(\delta)\mathcal{J}(u_{\mathcal{J}}^\dagger).$$

Therefore, we get an upper bound on $\mathcal{J}(u_\delta)$:

$$\mathcal{J}(u_\delta) \leqslant \frac{C(\delta)}{\alpha(\delta)} + \mathcal{J}(u_{\mathcal{J}}^\dagger). \tag{4.8}$$

The right-hand side is bounded uniformly in δ, since $\lim_{\delta\to 0} C(\delta)/\alpha(\delta) = 0$ by Assumption (iv) and $u_{\mathcal{J}}^\dagger$ does not depend on δ.

Choosing an arbitrary sequence $\delta_n \downarrow 0$, we conclude, using the $\tau_{\mathcal{U}}$-compactness of the sublevel sets of $\mathcal{J}$, that the sequence u_{δ_n} contains a $\tau_{\mathcal{U}}$-convergent subsequence (that we do not relabel to avoid triple subscripts) $u_{\delta_n} \stackrel{\tau_{\mathcal{U}}}{\to} u_0$. By Assumption (i) we get that $Au_{\delta_n} \stackrel{\tau_{\mathcal{V}}}{\to} Au_0$.

By Assumption (iii) $\mathcal{F}(\cdot,\cdot)$ is l.s.c. in both arguments and hence we get the following estimate

$$\begin{aligned}
\mathcal{F}(Au_0, f) &\leqslant \liminf_{n\to\infty} \mathcal{F}(Au_{\delta_n}, f_{\delta_n}) \leqslant \liminf_{n\to\infty} \mathcal{F}(Au_{\delta_n}, f_{\delta_n}) + \alpha(\delta_n)\mathcal{J}(u_{\delta_n}) \\
&\leqslant \liminf_{n\to\infty} \mathcal{F}(Au_{\mathcal{J}}^\dagger, f_{\delta_n}) + \alpha(\delta_n)\mathcal{J}(u_{\mathcal{J}}^\dagger) = \liminf_{n\to\infty} \mathcal{F}(f, f_{\delta_n}) + \alpha(\delta_n)\mathcal{J}(u_{\mathcal{J}}^\dagger) \\
&\leqslant \liminf_{n\to\infty} C(\delta_n) + \alpha(\delta_n)\mathcal{J}(u_{\mathcal{J}}^\dagger) = 0.
\end{aligned}$$

Here we used the facts that $\mathcal{J}(\cdot) \geqslant 0$ and u_{δ_n} is a minimiser in (4.6) with $\alpha = \alpha(\delta)$. Hence, u_0 solves problem (4.5).

Now it is left to show that u_0 has minimal value of $\mathcal{J}$ among all the minimisers of (4.5). Indeed, using Assumption (iv) and the fact that $\mathcal{J}(u_{\mathcal{J}}^\dagger) \leqslant \mathcal{J}(u)$ for any minimiser u of (4.5), we conclude from (4.8) that

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

Therefore, there exists $\lim_{n\to\infty} \mathcal{J}(u_{\delta_n}) = \mathcal{J}(u_0) = \mathcal{J}(u_{\mathcal{J}}^\dagger)$ and u_0 is a $\mathcal{J}$-minimising solution of (4.4).

□

Remark 4.2.5. The compactness of the level sets of $\mathcal{J}(u)$ in Assumption (ii) can be replaced by compactness of the level sets of $\Phi_f^\alpha(u) := \mathcal{F}(Au, f) + \alpha\mathcal{J}(u)$.

Remark 4.2.6. The theorem proves convergence of the regularised solutions in $\tau_{\mathcal{U}}$, which may differ from the strong topology. However, if $\mathcal{J}$ satisfies the *Radon-Riesz property* with respect to the topology $\tau_{\mathcal{U}}$, i.e. $u_j \stackrel{\tau_{\mathcal{U}}}{\to} u$ 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 $\tau_{\mathcal{U}}$ being the weak topology.

Examples of fidelity functions

Example 4.2.7. Let $\mathcal{V}$ be a Hilbert space and $\mathcal{F}(g,f) = \|g-f\|^2$. Obviously, in both arguments, it is strongly continuous (hence l.s.c.) and convex and, therefore, also weakly l.s.c (see Theorem 4.1.2). Therefore, Assumption (iii) of Theorem 4.2.4 is satisfied with $\tau_{\mathcal{V}}$ being the weak or the strong topology. Furthermore, the following properties hold

$$\mathcal{F}(f,f) = 0 \quad \text{and} \quad \mathcal{F}(f,f_\delta) \leqslant C(\delta) = \delta^2 \to 0 \quad \text{as } \delta \to 0$$

for all $f \in \mathcal{V}$ and all $f_\delta \in \mathcal{V}$ such that $\|f_\delta - f\| \leqslant \delta$. Therefore, with an appropriate choice of the regularisation functional $\mathcal{J}$, we obtain a convergent regularisation if

$$\lim_{\delta \to 0} \frac{\delta^2}{\alpha(\delta)} = 0.$$

(Compare this with the result obtained in (3.13) for Tikhonov regularisation).

Example 4.2.8. Let $\mathcal{V}$ be a Banach space and $\mathcal{F}(g,f) = \|g-f\|$. In both arguments, $\mathcal{F}(\cdot,\cdot)$ is strongly continuous (hence l.s.c.) and convex and, by Theorem 4.1.2, also weakly l.s.c. Assumption (iii) of Theorem 4.2.4 is again satisfied for the weak and the strong topologies. Furthermore, we see that

$$\mathcal{F}(f,f) = 0 \quad \text{and} \quad \mathcal{F}(f,f_\delta) \leqslant C(\delta) = \delta \to 0 \quad \text{as } \delta \to 0$$

for all $f \in \mathcal{V}$ and all $f_\delta \in \mathcal{V}$ such that $\|f_\delta - f\| \leqslant \delta$. Therefore, with an appropriate choice of the regularisation functional $\mathcal{J}$, we obtain a convergent regularisation if

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

A typical choice is $\mathcal{V} = L^1(\Omega)$ for some bounded set $\Omega \subset \mathbb{R}^d$ and $\mathcal{F}(g,f) = \|g-f\|_1$.

Examples of regularisers

Example 4.2.9. Let $\mathcal{U}$ be a Hilbert space and $\mathcal{J}(u) = \|u\|^2$. As discussed in Example 4.2.7, the norm in a Hilbert space is weakly l.s.c. By Theorem 4.1.2 we know that (norm) bounded sequences have weakly convergent subsequences. Therefore, Assumption (ii) of Theorem 4.2.4 is satisfied with $\tau_{\mathcal{U}}$ 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 $W^{1,2}$ 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 $\mathcal{J}(u) = \|u\|_{W^{1,2}}$ 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 4.2.10 (ℓ^1-regularisation)**.** Let $\mathcal{U} = \ell^2$ 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 4.1.5 $\mathcal{J}(\cdot)$ is weakly l.s.c. in ℓ^2. It is evident that $\ell^p \subset \ell^q$ and $\|\cdot\|_{\ell^p} \leqslant \|\cdot\|_{\ell^q}$ for $p \geqslant q$; in particular, we have that $\ell^2 \subset \ell^1$ and $\|\cdot\|_{\ell^2} \leqslant \|\cdot\|_{\ell^1}$. 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 4.2.4 is satisfied with $\tau_{\mathcal{U}}$ being the weak topology in ℓ^2. Hence, we get weak convergence of regularised solutions in ℓ^2 and using the Radon-Riesz property we conclude that the convergence is actually strong.

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 $\operatorname{supp}(u) = \{i \in \mathbb{N} \,|\, u_i \neq 0\}$. 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

$$\min_{u \in \ell^2} \left\{ \frac{1}{2} \|Au - f\|_{\ell^2}^2 + \alpha \|u\|_1 \right\}. \tag{4.9}$$

is also called *lasso* in the statistical literature.

Note that, although both $\mathcal{J}(u) = \|u\|_{\ell^2}$ and $\mathcal{J}(u) = \|u\|_{\ell^1}$ give the same type of convergence – strong convergence in ℓ^2 (i.e., their topological properties are the same), the structure of the regularised solutions is quite different. Therefore, topological considerations should not be the only ones when choosing an appropriate regularisation for an applied problem. Studying the geometric structure of the regularised solutions is an important aspect of modern inverse imaging problems.

4.3 Total Variation Regularisation

As pointed out in Example 4.2.9, in imaging we are interested in regularisers that allow for discontinuities while maintaining sufficient regularity of the reconstructions. One very popular choice is the so-called *total variation* regulariser.

Definition 4.3.1. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and $u \in L^1(\Omega)$. Let $\mathcal{D}(\Omega, \mathbb{R}^n)$ be the following set of vector-valued* test functions *(i.e. functions that map from Ω to $\mathbb{R}^n$)*

$$\mathcal{D}(\Omega, \mathbb{R}^n) := \left\{ \varphi \in C_0^\infty(\Omega; \mathbb{R}^n) \;\middle|\; \operatorname{ess\,sup}_{x \in \Omega} \|\varphi(x)\|_2 \leqslant 1 \right\}.$$

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

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

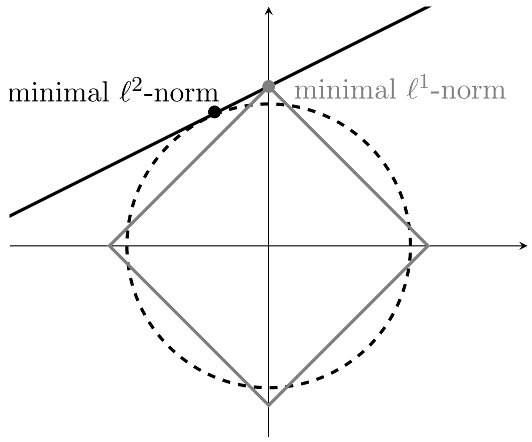

Figure 4.8: Non-injective operators have a non-trivial kernel such that the inverse problem has more than one solution and the solutions form an affine subspace visualised by the solid line. Different regularisation functionals favour different solutions. The circle and the diamond indicate all points 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.*

Remark 4.3.2. Definition 4.3.1 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 $L^1(\Omega, \mathbb{R}^n)$ (i.e. u belongs to the Sobolev space $W^{1,1}(\Omega)$) we obtain, integrating by parts, that

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

By the Cauchy-Schwartz 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 $\mathcal{D}(\Omega, \mathbb{R}^n)$, but we can approximate it with functions from $\mathcal{D}(\Omega, \mathbb{R}^n)$, since any function in $W^{1,1}(\Omega)$ can be approximated with smooth functions [2, Thm. 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 $\varphi \in \mathcal{D}(\Omega, \mathbb{R}^n)$ is equal to

$$\mathrm{TV}(u) = \int_\Omega \|\nabla u(x)\|_2 \, dx.$$

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.

Proposition 4.3.3. TV *is a proper and convex functional* $L^1(\Omega) \to \bar{\mathbb{R}}$. *For any constant function* $\mathbf{c}\colon \mathbf{c}(x) \equiv c \in \mathbb{R}$ *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. □

Definition 4.3.4. *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*-spcae) defined as follows*

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

It can be shown that BV is a Banach space [4].

Remark 4.3.5. The definition of total variation (Def. 4.3.1) looks much like that of a dual norm in some Banach space. Indeed, it can be shown that TV is a norm on a subspace of BV and that this space is a dual of some Banach space.

Example 4.3.6 (TV of an indicator function)**.** Suppose $\mathcal{C} \subset \Omega \subset \mathbb{R}^2$ is a bounded domain with smooth boundary and $u(\cdot) = \mathbf{1}_{\mathcal{C}}(\cdot)$ is its indicator function, i.e.

$$\mathbf{1}_{\mathcal{C}}(u) = \begin{cases} 1 & u \in \mathcal{C} \\ 0 & u \in \mathcal{U} \setminus \mathcal{C} \end{cases}.$$

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

$$\int_\Omega u(x) \operatorname{div} \varphi(x)\, dx = \int_{\mathcal{C}} \operatorname{div} \varphi(x)\, dx = \int_{\partial \mathcal{C}} \langle \varphi(x), \mathbf{n}_{\partial \mathcal{C}}(x) \rangle \, dl,$$

where $\partial \mathcal{C}$ is the boundary of $\mathcal{C}$ and $\mathbf{n}_{\partial \mathcal{C}}(x)$ is the unit normal at x. We, obviously, have that for every x

$$\langle \varphi(x), \mathbf{n}(x) \rangle = \frac{1}{2}(\|\varphi(x)\|^2 + \|\mathbf{n}_{\partial \mathcal{C}}(x)\|^2 - \|\varphi(x) - \mathbf{n}_{\partial \mathcal{C}}(x)\|^2),$$

so we get that

$$\mathrm{TV}(u) = \sup_{\varphi \in \mathcal{D}(\Omega, \mathbb{R}^n)} \int_{\partial \mathcal{C}} \frac{1}{2}(\|\varphi(x)\|^2 + \|\mathbf{n}_{\partial \mathcal{C}}(x)\|^2 - \|\varphi(x) - \mathbf{n}_{\partial \mathcal{C}}(x)\|^2)\, dl.$$

Since $\partial \mathcal{C}$ is smooth and $\|\mathbf{n}_{\partial \mathcal{C}}(x)\| = 1$ for every x, $\mathbf{n}_{\partial \mathcal{C}}$ can be extended to feasible vector field on Ω (i.e. one that is in $D(\Omega, \mathbb{R}^n)$) and the supremum is attained at $\varphi = \mathbf{n}_{\partial \mathcal{C}}$. Therefore, we get that

$$\mathrm{TV}(u) = \int_{\partial \mathcal{C}} \|\mathbf{n}_{\partial \mathcal{C}}(x)\|^2 \, dl = \int_{\partial \mathcal{C}} 1 \cdot dl = \mathrm{Per}(\mathcal{C}),$$

where $\mathrm{Per}(\mathcal{C})$ is the perimeter of $\mathcal{C}$.

Therefore, total variation of the characteristic function of a domain with smooth boundary is equal to its perimeter. This can be extended to domains with Lipschitz boundary by constructing a sequence of functions in $D(\Omega, \mathbb{R}^n)$ that converge pointwise to $\mathbf{n}_{\partial \mathcal{C}}$.

To apply Theorem 4.2.4, we need to study the properties of TV as a functional $L^1(\Omega) \to \bar{\mathbb{R}}$. First of all, we note that $\mathrm{BV}(\Omega)$ is compactly embedded in $L^1(\Omega)$. We start with the following classical result.

Theorem 4.3.7 (Rellich-Kondrachov, [2, Thm. 6.3])**.** *Let $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain (i.e. non-empty, open, connected and bounded with Lipschitz boundary) and either*

$$\begin{array}{rlll} & n > mp & \textit{and} & p^* := np/(n - mp) \\ \textit{or} & n \leqslant mp & \textit{and} & p^* := \infty\,. \end{array}$$

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

Since functions from $\mathrm{BV}(\Omega)$ can be approximated by smooth functions [4, Thm. 3.9], the Rellich-Kandrachov Theorem (for $m = 1, p = 1$) gives us compactness for $\mathrm{BV}(\Omega)$.

Corollary 4.3.8 ([4, Corrollary 3.49]). *For any bounded Lipschitz domain $\Omega \subset \mathbb{R}^n$ the embedding*

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

is compact.

Therefore, the level sets of $\mathcal{J}(u) = \|u\|_{\mathrm{BV}}$ are strongly sequentially compact in $L^1(\Omega)$. This is one of the ingredients we need to apply Theorem 4.2.4. The other one is lower-semicontinuity, which is guaranteed by the following theorem.

Theorem 4.3.9. *Let $\Omega \subset \mathbb{R}^n$ be open and bounded. Then the total variation is strongly l.s.c. in $L^1(\Omega)$.*

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

$$\int_\Omega [u(x) - u_j(x)] \operatorname{div} \varphi(x) dx \leqslant \underbrace{\int_\Omega |u(x) - u_j(x)| dx}_{=\|u-u_j\|_{L^1} \to 0} \; \underbrace{\operatorname{ess\,sup}_{x\in\Omega} |\operatorname{div} \varphi(x)|}_{<\infty} \to 0$$

and therefore

$$\begin{aligned} \int_\Omega u(x) \operatorname{div} \varphi(x) dx &= \lim_{j\to\infty} \int_\Omega u_j(x) \operatorname{div} \varphi(x) dx = \liminf_{j\to\infty} \int_\Omega u_j(x) \operatorname{div} \varphi(x) dx \\ &\leqslant \liminf_{j\to\infty} \sup_{\varphi\in\mathcal{D}(\Omega,\mathbb{R}^n)} \int_\Omega u_j(x) \operatorname{div} \varphi(x) dx \leqslant \liminf_{j\to\infty} \mathrm{TV}(u_j). \end{aligned}$$

Taking the supremum over all test functions on the left-hand side (and noting that the right-hand side already does not depend on φ), we get the assertion:

$$\mathrm{TV}(u) = \sup_{\varphi\in\mathcal{D}(\Omega,\mathbb{R}^n)} \int_\Omega u(x) \operatorname{div} \varphi(x) dx \leqslant \liminf_{j\to\infty} \mathrm{TV}(u_j).$$

Note that the left and right hand sides may not be finite. □

This result can actually be strengthened, as the following corollary shows.

Corollary 4.3.10. *Let $\Omega \subset \mathbb{R}^n$ be open and bounded. Then the total variation is both weakly and strongly l.s.c. in any $L^p(\Omega)$ space with $1 \leqslant p < \infty$.*

Proof. Since Ω is bounded, $\|u\|_1 \leqslant |\Omega|^{1-1/p} \|u\|_p$ for any $p > 1$ (where $|\Omega| := \int_\Omega 1 \cdot dx$) and strong convergence in $L^p(\Omega)$ with $p > 1$ implies strong convergence in $L^1(\Omega)$. Strong lower-semicontinuity of the total variation in $L^p(\Omega)$ with $p > 1$ then follows from strong lower-semicontinuity in $L^1(\Omega)$.

Weak lower-semicontinuity in $L^1(\Omega)$ follows from the fact that stong and weak sequential convergences coincide in $L^1(\Omega)$ [11, Corollary IV.8.14].

As pointed out in Remark 4.1.38, in reflexive spaces strongly lower-semicontinuous convex functions are also weakly lower-semicontinuous. Since the spaces $L^p(\Omega)$ with $1 < p < \infty$ are reflexive and by Proposition 4.3.3 total variation is convex, we get weak lower semicontinuity in $L^p(\Omega)$ for any $1 < p < \infty$. □

Remark 4.3.11. Combining these results, we conclude that with a suitable fidelity function and a suitable parameter choice rule the regulariser $\mathcal{J}(u) = \mathrm{TV}(u) + \|u\|_1$ ensures strong L^1-convergence of the regularised solutions. If the forward operator is such that boundedness of the fidelity term implies boundedness of $\|u\|_1$, then the term $\|u\|_1$ can be dropped and $\mathcal{J}(u) = \mathrm{TV}(u)$ can be used instead, ensuring the same convergence properties. See Remark 4.3.14 for an example of a situation when this is the case.

If $\Omega \subset \mathbb{R}^2$, it is often useful to consider TV not as a functional in $L^1(\Omega)$, but as a functional in $L^2(\Omega)$. By Corollary 4.3.10, weak lower-semicontinuity holds in this case. However, the compact embedding result in Theorem 4.3.8 does not hold any more (in two dimensions, the embedding $\mathrm{BV}(\Omega) \to L^2(\Omega)$ is continuous, but not compact [4, Corrollary 3.49]). However, the following result helps.

Proposition 4.3.12 ([4, Remark 3.50]). *Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain. Then there exists a constant $C > 0$ such that for all $u \in \mathrm{BV}(\Omega)$ the* Poincaré–Wirtinger *type inequality is satisfied*

$$\|u - u_\Omega\|_{L^2} \leqslant C\,\mathrm{TV}(u),$$

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

Corollary 4.3.13. *It is often useful to consider a subspace $\mathrm{BV}_0(\Omega) \subset \mathrm{BV}(\Omega)$ of functions with zero mean, i.e.*

$$\mathrm{BV}_0(\Omega) := \{u \in \mathrm{BV}(\Omega) \colon \int_\Omega u(x)dx = 0\}. \tag{4.10}$$

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

$$\|u\|_{L^2} \leqslant C\,\mathrm{TV}(u).$$

Remark 4.3.14. Total variation is often used in conjunction with the L^2-fidelity (see Example 4.2.7) under the assumption that the forward operator $A\colon L^2(\Omega) \to L^2(\Omega)$ satisfies $A\mathbf{1} \neq 0$, where $\mathbf{1}(x) \equiv 1$ for all x. In this case, the boundedness of $\mathrm{TV}(u)$ together with the boundedness of the fidelity term $\|Au - f_\delta\|_2^2$ imply the boundedness of the mean value $u_\Omega \in \mathbb{R}$.

Indeed, suppose that there exists a sequence u^n is such that u_Ω^n is unbounded. Then, since $A\mathbf{1} \neq 0$, the sequence Au_Ω^n is also unbounded. Consider $u_0^n := u^n - u_\Omega^n \in \mathrm{BV}_0(\Omega)$. By Proposition 4.3.3 we have that

$$\mathrm{TV}(u_0^n) = \mathrm{TV}(u^n - u_\Omega^n) = \mathrm{TV}(u^n)$$

and therefore bounded. We also have that

$$\begin{aligned}\|Au_\Omega^n\|_2 &= \|Au_\Omega^n + Au_0^n - f_\delta - (Au_0^n - f_\delta)\|_2 \leqslant \|Au^n - f_\delta\|_2 + \|Au_0^n - f_\delta\|_2 \\ &\leqslant \|Au^n - f_\delta\|_2 + \|A\|_{L^2\to L^2}\|u_0^n\|_2 + \|f_\delta\|_2.\end{aligned}$$

The first term on the right-hand side is bounded by assumption; the second one is bounded, since the bound on $\mathrm{TV}(u_0^n)$ provides a bound on $\|u_0^n\|_2$ and A is bounded; the third one is bounded, since $\|f_\delta - f\|_2 \to 0$. Therefore, $\|Au_\Omega^n\|_2$ is bounded, which is a contradiction.

These derivations show that using a combination of $\mathcal{F}(Au, f_\delta) = \|Au - f_\delta\|_2^2$ and $\mathcal{J}(u) = \mathrm{TV}(u)$ under the assumption that $A\mathbf{1} \neq 0$ guarantees that the L^2-norm $\|u_\delta\|_2$ of the regularised solution u_δ is bounded uniformly in δ. Since bounded sequences in L^2 have weakly convergent subsequences (Theorem 4.1.2), this guarantees weak L^2-convergence of u_δ as $\delta \to 0$ by Theorem 4.2.4. Furthermore, since $\|u\|_1 \leqslant |\Omega|^{1/2}\|u\|_2$ for any $u \in L^2(\Omega)$, we get a bound on $\|u_\delta\|_1$ and hence strong convergence in $L^1(\Omega)$ by Remark 4.3.11.

This setting is widely used in imaging applications [23]. For instance, the so-called ROF model for image denoising [21] consists in minimising the following functional

$$\min_{u \in L^2(\Omega)} \|u - f_\delta\|_2^2 + \alpha\,\mathrm{TV}(u). \tag{4.11}$$

In this case, the forward operator is the identity operator (and $A\mathbf{1} \neq 0$ is satisfied trivially). More generally, one considers the following optimisation problem

$$\min_{u \in L^2(\Omega)} \|Au - f_\delta\|_2^2 + \alpha \operatorname{TV}(u), \tag{4.12}$$

where $A\colon L^2(\Omega) \to L^2(\Omega)$ is injective. Injectiveness is equivalent to the condition that the null-space is a singleton, i.e. $\{u \in L^2(\Omega)\colon Au = 0\} = \{0\}$), and guarantees that $A\mathbf{1} \neq 0$.

Chapter 5

Dual Perspective

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 distances* associated with the (convex) regularisation functional $\mathcal{J}$. Recall that for a convex functional $\mathcal{J}$, $u, v \in \mathcal{U}$ such that $\mathcal{J}(v) < \infty$ and $q \in \partial\mathcal{J}(v)$, the (generalised) Bregman distance is given by the following expression (cf. Def. 4.1.23)

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

Also widely used is the *symmetric* Bregman distance (cf. Def. 4.1.25) given by the following expression (here $p \in \partial\mathcal{J}(u)$)

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

Bregman distances 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 L^2- regularisation with $\mathcal{J}(u) = \frac{1}{2}\|u\|^2_\mathcal{U}$, 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

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

which happens to coincide with the symmetric Bregman distance. Therefore, in the classical L^2-case, the Bregman distance just measures the L^2-distance between a regularised solution and a $\mathcal{J}$-minimising solution.

We are looking for a convergence rate of the following form

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

where $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a known function of δ such that $\psi(\delta) \to 0$ as $\delta \to 0$. To obtain such an estimate, we need to not only understand the convergence of u_δ (to $u^\dagger_\mathcal{J}$), but also that of the subgradient $p_\delta \in \partial\mathcal{J}(u_\delta)$, which should ideally converge to some $p_\mathcal{J} \in \partial\mathcal{J}(u^\dagger_\mathcal{J})$.

5.1 Dual Problem

Recall that u_δ solves the following problem

$$\min_{u\in\mathcal{U}} \mathcal{F}(Au, f_\delta) + \alpha\mathcal{J}(u) \tag{5.1}$$

with an appropriately chosen $\alpha = \alpha(\delta)$. In this Chapter we will assume that $\mathcal{U}$ and $\mathcal{V}$ are Hilbert spaces and will choose the standard fidelity $\mathcal{F}(Au, f) = \frac{1}{2}\|Au - f\|_{\mathcal{V}}^2$. We will also assume that the regulariser is proper, convex, l.s.c., absolute one-homogeneous and satisfies conditions of Theorem 4.2.4.

In this case problem (5.1) takes the following form

$$\min_{u\in\mathcal{U}} \frac{1}{2}\|Au - f_\delta\|_{\mathcal{V}}^2 + \alpha\mathcal{J}(u). \tag{5.2}$$

We will see that all subgradients $p_\delta \in \partial\mathcal{J}(u_\delta)$ are closely related to solutions of the dual problem of (5.2) in the sense of duality in convex optimisation [13].

Let us drop the subscript δ for some time and consider the function $\varphi\colon \mathcal{V} \to \mathbb{R}$, $\varphi(x) := \frac{1}{2}\|x - f\|_{\mathcal{V}}^2$, where $f \in \mathcal{V}$ is a parameter. The Fenchel conjugate of φ is given by

$$\varphi^*(\nu) = \sup_{x\in\mathcal{V}} \langle \nu, x\rangle - \varphi(x) = \sup_{x\in\mathcal{V}} \langle \nu, x\rangle - \frac{1}{2}\|x - f\|_{\mathcal{V}}^2, \quad \nu \in \mathcal{V}.$$

The supremum is attained at $x = \nu + f$ and therefore

$$\varphi^*(\nu) = \langle \nu, \nu + f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2 = \langle \nu, f\rangle + \frac{1}{2}\|\nu\|_{\mathcal{V}}^2.$$

By Theorem 4.1.16 we have that φ is equal to its biconjugate, i.e.

$$\varphi(x) = \sup_{\nu\in\mathcal{V}} \langle \nu, x\rangle - \varphi^*(\nu) = \sup_{\nu\in\mathcal{V}} \langle \nu, x\rangle - \langle \nu, f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2 = \sup_{\nu\in\mathcal{V}} \langle \nu, x - f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2.$$

For $x = Au$, therefore, we get that

$$\varphi(Au) = \frac{1}{2}\|Au - f\|_{\mathcal{V}}^2 = \sup_{\nu\in\mathcal{V}} \langle \nu, Au - f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2.$$

Obviously, the objective function attains its maximum (at $\nu = Au - f$), so we can replace the supremum with a maximum. Note that we took the operator outside the norm, which can be useful in numerical optimisation algorithms [10].

Now we can rewrite (5.2) as follows

$$\min_{u\in\mathcal{U}} \max_{\nu\in\mathcal{V}} \langle \nu, Au - f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2 + \alpha\mathcal{J}(u). \tag{5.3}$$

Problem (5.3) is called the *saddle-point* problem. If it has a solution then we can easily derive optimality conditions by differentiating the objective function in u and ν:

$$\nu = Au - f, \quad A^*\left(\frac{-\nu}{\alpha}\right) \in \partial\mathcal{J}(u). \tag{5.4}$$

We can swap the minimum and the maximum in (5.3) under the conditions given in [13, Ch.III Thm 4.1 and Rem. 4.2], i.e. that $\varphi(x)$ is continuous at $x = 0$, $\varphi(0) < +\infty$ and $\mathcal{J}(0) < +\infty$ (these conditions are, obviously, satisfied). We get

$$\begin{aligned}\min_{u\in\mathcal{U}} \max_{\nu\in\mathcal{V}} \langle \nu, Au - f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2 + \alpha\mathcal{J}(u) &= \max_{\nu\in\mathcal{V}} \min_{u\in\mathcal{U}} \langle \nu, Au - f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2 + \alpha\mathcal{J}(u) \\ &= \max_{\nu\in\mathcal{V}} \left\{ \left[\min_{u\in\mathcal{U}} \langle \nu, Au\rangle + \alpha\mathcal{J}(u)\right] - \langle \nu, f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2 \right\}. \end{aligned} \tag{5.5}$$

The minimum of the expression in the square brackets is given by

$$\begin{aligned}\min_{u\in\mathcal{U}} \langle \nu, Au\rangle + \alpha\mathcal{J}(u) &= \min_{u\in\mathcal{U}} \langle A^*\nu, u\rangle + \alpha\mathcal{J}(u)\\ &= -\alpha \max_{u\in\mathcal{U}} \left\langle A^*\left(\frac{-\nu}{\alpha}\right), u\right\rangle - \mathcal{J}(u) = -\alpha\mathcal{J}^*\left(A^*\left(\frac{-\nu}{\alpha}\right)\right).\end{aligned}$$

Since $\mathcal{J}$ is absolute one-homogeneous, its Fenchel conjugate is the characteristic function of $\partial\mathcal{J}(0)$ (Prop. 4.1.29) and we get

$$\min_{u\in\mathcal{U}} \langle \nu, Au\rangle + \alpha\mathcal{J}(u) = -\alpha\chi_{\partial\mathcal{J}(0)}\left(A^*\left(\frac{-\nu}{\alpha}\right)\right).$$

Substituting this into (5.5), we get

$$\max_{\nu\in\mathcal{V}}\left\{\left[\min_{u\in\mathcal{U}} \langle \nu, Au\rangle + \alpha\mathcal{J}(u)\right] - \langle \nu, f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2\right\} = \max_{\nu\in\mathcal{V}\colon A^*\left(\frac{-\nu}{\alpha}\right)\in\partial\mathcal{J}(0)} - \langle \nu, f\rangle - \frac{1}{2}\|\nu\|_{\mathcal{V}}^2.$$

Denoting $\mu := -\frac{\nu}{\alpha} \in \mathcal{V}$, we rewrite this problem as follows

$$\max_{\mu\in\mathcal{V}\colon A^*\mu\in\partial\mathcal{J}(0)} \alpha\left(\langle \mu, f\rangle - \frac{\alpha}{2}\|\mu\|_{\mathcal{V}}^2\right). \tag{5.6}$$

Problem (5.6) is called the *dual problem*. With this notation, optimality conditions (5.4) take the following form

$$A^*\mu \in \partial\mathcal{J}(u), \quad \mu = \frac{f - Au}{\alpha}. \tag{5.7}$$

It can be shown [13] that for any feasible solution u_0 of the primal problem (5.2) and for any feasible solution μ_0 of the dual problem (5.6), the objective value of the dual problem does not exceed that of the primal problem, i.e.

$$\frac{1}{2}\|Au_0 - f\|_{\mathcal{V}}^2 + \alpha\mathcal{J}(u_0) \geqslant \alpha\langle \mu_0, f\rangle - \frac{\alpha^2}{2}\|\mu_0\|_{\mathcal{V}}^2. \tag{5.8}$$

This also holds for the optimal solutions u_δ and μ_δ (where we return to the notation with δ). The difference

$$\frac{1}{2}\|Au_\delta - f_\delta\|_{\mathcal{V}}^2 + \alpha\mathcal{J}(u_\delta) - \left(\alpha\langle \mu_\delta, f_\delta\rangle - \frac{\alpha^2}{2}\|\mu_\delta\|_{\mathcal{V}}^2\right) \geqslant 0 \tag{5.9}$$

is referred to as the *duality gap*. The fact that it is always non-negative is referred to as *weak duality*. Under some assumptions (for instance, those in Theorem 5.1.1) the duality gap is zero; in this case it is said that *strong duality* holds.

Existence of a solution is guaranteed by the following Theorem.

Theorem 5.1.1 ([13, Ch.III Thm 4.1 and Rem. 4.2]). *Consider the primal problem* (5.2) *in the general from*

$$\min_{u\in\mathcal{U}} E(Au) + F(u),$$

where $E\colon \mathcal{V} \to \bar{\mathbb{R}}$ *and* $F\colon \mathcal{U} \to \bar{\mathbb{R}}$. *Suppose that*

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

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

Then

(i) Both the primal problem and its dual have solutions, which we denote by $\hat{u}$ *and* $\hat{\eta}$*, respectively;*

(ii) There is no duality gap between the primal and the dual problems, i.e. strong duality holds;

(iii) The following optimality conditions hold

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

In our case, $E(Au) = \frac{1}{2}\|Au - f\|_{\mathcal{V}}^2$ and $F(u) = \alpha\mathcal{J}(u)$. Condition (i) is satisfied by the assumptions of Theorem 4.2.4 (in particular, coercivity is implied by the compactness of the sub-level sets). Condition (ii) is satisfied at $u_0 = 0$. Therefore, for any $\delta > 0$ there exists a solution u_δ of the primal problem (5.2) and μ_δ of the dual problem (5.6) and by strong duality we have that

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

Optimality conditions (iii) in this case take the following form (cf. (5.7))

$$A^*\mu_\delta \in \partial\mathcal{J}(u_\delta), \quad \mu_\delta = \frac{f_\delta - Au_\delta}{\alpha(\delta)}, \tag{5.10}$$

where we remind ourselves of the fact that α is chosen according to a parameter choice rule $\alpha(\delta)$.

5.2 Source Condition

Formal limits of problems (5.2) and (5.6) at $\delta = 0$ are

$$\inf_{u\colon Au=f} \mathcal{J}(u) = \inf_{u\in\mathcal{U}} \chi_{\{f\}}(Au) + \mathcal{J}(u) \tag{5.11}$$

and

$$\begin{aligned}\sup_{\mu\colon A^*\mu\in\partial\mathcal{J}(0)} \left\langle \mu, f \right\rangle &= \sup_{\mu\colon A^*\mu\in\partial\mathcal{J}(0)} \left\langle \mu, Au_{\mathcal{J}}^\dagger \right\rangle \\ &= \sup_{\mu\colon A^*\mu\in\partial\mathcal{J}(0)} \left\langle A^*\mu, u_{\mathcal{J}}^\dagger \right\rangle = \sup_{v\in\mathcal{R}(A^*)\cap\partial\mathcal{J}(0)} \left\langle v, u_{\mathcal{J}}^\dagger \right\rangle. \end{aligned} \tag{5.12}$$

Since the characteristic function $\chi_{\{f\}}(\cdot)$ is not continuous anywhere in its domain, Theorem 5.1.1 does not apply and we cannot expect strong duality in general. We cannot even guarantee that a solution of the dual limit problem (5.12) exists. Indeed, the feasible set $\mathcal{R}(A^*) \cap \partial\mathcal{J}(0)$ in (5.12) may be empty and even if it is not a solution may not exist, since $\mathcal{R}(A^*)$ is not closed (strongly and hence weakly, since it is convex [11, Thm. V.3.13]).

Therefore, the behaviour of μ_δ as $\delta \to 0$ is unclear. (Recall that we need to understand this behaviour to obtain an estimate on the Bregman distance between u_δ and $u_{\mathcal{J}}^\dagger$). A natural question to ask is whether μ_δ remains bounded as $\delta \to 0$.

Theorem 5.2.1 (Necessary conditions, [18])**.** *Suppose that conditions of Theorem 4.2.4 are satisfied with* $\tau_{\mathcal{U}}$ *and* $\tau_{\mathcal{V}}$ *being weak topologies in* $\mathcal{U}$ *and* $\mathcal{V}$*, respectively. Suppose that* μ_δ *is bounded uniformly in* δ*. Then there exists* $\mu^\dagger \in \mathcal{V}$ *such that* $A^*\mu^\dagger \in \partial\mathcal{J}(u_{\mathcal{J}}^\dagger)$*.*

Definition 5.2.2 (Source condition [9]). *We say that a $\mathcal{J}$-minimising solution $u^\dagger_\mathcal{J}$ satisfies the* source condition *if $\exists \mu^\dagger \in \mathcal{V}$ such that $A^*\mu^\dagger \in \partial\mathcal{J}(u^\dagger_\mathcal{J})$, i.e. if*

$$\mathcal{R}(A^*) \cap \partial\mathcal{J}(u^\dagger_\mathcal{J}) \neq \emptyset. \tag{5.13}$$

Proof of Theorem 5.2.1. Consider an arbitrary sequence $\delta_n \downarrow 0$. Since $\|\mu_\delta\|_\mathcal{V} \leqslant C$ for all δ, by weak compactness of a ball in a Hilbert space we get that there exists a weakly convergent subsequence (that we do not relabel), i.e.

$$\mu_{\delta_n} \rightharpoonup \mu_0 \in \mathcal{V}.$$

By the weak-weak continuity of A^* we get that

$$A^*\mu_{\delta_n} \rightharpoonup A^*\mu_0.$$

Since $\partial\mathcal{J}(0)$ is weakly closed (Theorem 4.1.19) and $A^*\mu_{\delta_n} \in \partial\mathcal{J}(0)$ (see optimality conditions (5.10), we get that

$$A^*\mu_0 \in \partial\mathcal{J}(0).$$

Since $\mathcal{J}$ is absolute one-homogeneous, we get by Proposition 4.1.27 that

$$\langle A^*\mu_{\delta_n}, u_{\delta_n}\rangle = \mathcal{J}(u_{\delta_n}).$$

We also note that

$$\begin{aligned}\langle A^*\mu_{\delta_n}, u_{\delta_n}\rangle &= \left\langle A^*\mu_{\delta_n}, u^\dagger_\mathcal{J}\right\rangle + \left\langle A^*\mu_{\delta_n}, u_{\delta_n} - u^\dagger_\mathcal{J}\right\rangle \\ &= \left\langle A^*\mu_{\delta_n}, u^\dagger_\mathcal{J}\right\rangle + \langle \mu_{\delta_n}, Au_{\delta_n} - f\rangle \leqslant \left\langle A^*\mu_{\delta_n}, u^\dagger_\mathcal{J}\right\rangle + \|\mu_{\delta_n}\|_\mathcal{V}\|Au_{\delta_n} - f\|_\mathcal{V}.\end{aligned}$$

Since $\|\mu_{\delta_n}\|_\mathcal{V}$ is bounded and $\|Au_{\delta_n} - f\|_\mathcal{V} \to 0$, we get that

$$\langle A^*\mu_{\delta_n}, u_{\delta_n}\rangle \to \left\langle A^*\mu_0, u^\dagger_\mathcal{J}\right\rangle.$$

On the other hand, we know that $\mathcal{J}(u_{\delta_n}) \to \mathcal{J}(u^\dagger_\mathcal{J})$. Therefore, we get that

$$\mathcal{J}(u^\dagger_\mathcal{J}) = \left\langle A^*\mu_0, u^\dagger_\mathcal{J}\right\rangle.$$

Since $A^*\mu_0 \in \partial\mathcal{J}(0)$ and $\mathcal{J}(u^\dagger_\mathcal{J}) = \left\langle A^*\mu_0, u^\dagger_\mathcal{J}\right\rangle$, we conclude, using Proposition 4.1.30, that $A^*\mu_0 \in \partial\mathcal{J}(u^\dagger_\mathcal{J})$ and the assertion of the Theorem holds with $\mu^\dagger = \mu_0$. □

So, the source condition is necessary for the boundedness of μ_δ. It turns out to be also sufficient.

Theorem 5.2.3 (Sufficient conditions, [18]). *Suppose that the source condition* (5.13) *is satisfied at a $\mathcal{J}$-minimising solution $u^\dagger_\mathcal{J}$ and suppose that $\alpha(\delta)$ is chosen such that $\frac{\delta}{\alpha(\delta)}$ is uniformly bounded. Then μ_δ is bounded uniformly in δ. Moreover, $\mu_\delta \to \mu^\dagger$ strongly in $\mathcal{V}$ as $\delta \to 0$, where $\mu^\dagger$ is the solution of the dual limit problem* (5.12) *with minimal norm.*

Proof. The source condition (5.13) guarantees that $\exists \mu_0 \in \mathcal{V}$ s.t. $A^*\mu_0 \in \partial\mathcal{J}(u^\dagger_\mathcal{J})$, i.e. that

$$\begin{cases} A^*\mu_0 \in \partial\mathcal{J}(0), \\ \mathcal{J}(u^\dagger_\mathcal{J}) = \left\langle A^*\mu_0, u^\dagger_\mathcal{J}\right\rangle = \left\langle \mu_0, Au^\dagger_\mathcal{J}\right\rangle = \langle \mu_0, f\rangle. \end{cases}$$

For any feasible solution μ of the dual limit problem (5.12) we have that

$$\langle \mu, f \rangle \leqslant \mathcal{J}(u^\dagger_{\mathcal{J}}),$$

since weak duality between the limit primal and dual problems holds (problems (5.11) and (5.12), respectively) and $u^\dagger_{\mathcal{J}}$ is a feasible solution of (5.11). Therefore, μ_0 solves the dual limit problem (5.12) and

$$\langle \mu_0, f \rangle \geqslant \langle \mu_\delta, f \rangle \quad \forall \delta, \tag{5.14}$$

since μ_δ is feasible in (5.12).

Analogously, since μ_δ solves the dual problem (5.6) and μ_0 is feasible in (5.6), we get that for all δ

$$\langle \mu_\delta, f_\delta \rangle - \frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{V}} \geqslant \langle \mu_0, f_\delta \rangle - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{V}}. \tag{5.15}$$

Therefore,

$$\begin{aligned} \frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{V}} - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{V}} \leqslant \langle \mu_\delta, f_\delta \rangle - \langle \mu_0, f_\delta \rangle &\leqslant \langle \mu_\delta, f_\delta - f \rangle + \langle \mu_\delta, f \rangle \\ &- \langle \mu_0, f \rangle + \langle \mu_0, f - f_\delta \rangle \leqslant \langle \mu_0 - \mu_\delta, f - f_\delta \rangle \\ &\leqslant \delta\|\mu_0 - \mu_\delta\|_{\mathcal{V}} \leqslant \delta(\|\mu_0\|_{\mathcal{V}} + \|\mu_\delta\|_{\mathcal{V}}). \end{aligned}$$

Noting that

$$\frac{\alpha}{2}\|\mu_\delta\|^2_{\mathcal{V}} - \frac{\alpha}{2}\|\mu_0\|^2_{\mathcal{V}} = \frac{\alpha}{2}(\|\mu_\delta\|_{\mathcal{V}} - \|\mu_0\|_{\mathcal{V}})(\|\mu_\delta\|_{\mathcal{V}} + \|\mu_0\|_{\mathcal{V}}),$$

we get that

$$\frac{\alpha}{2}(\|\mu_\delta\|_{\mathcal{V}} - \|\mu_0\|_{\mathcal{V}})(\|\mu_0\|_{\mathcal{V}} + \|\mu_\delta\|_{\mathcal{V}}) \leqslant \delta\|\mu_0 - \mu_\delta\|_{\mathcal{V}} \leqslant \delta(\|\mu_0\|_{\mathcal{V}} + \|\mu_\delta\|_{\mathcal{V}})$$

and

$$\|\mu_\delta\|_{\mathcal{V}} \leqslant \|\mu_0\|_{\mathcal{V}} + \frac{2\delta}{\alpha} \leqslant C, \tag{5.16}$$

since $\frac{\delta}{\alpha}$ is bounded.

By weak compactness of a ball in a Hilbert space, we conclude that for any sequence $\delta_n \downarrow 0$ there exists a subsequence (which we do not relabel) such that

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

By weak-weak continuity of A^* and weak closedness of $\partial\mathcal{J}(0)$ (Theorem 4.1.19) we get that

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

and μ^* is feasible in (5.12). Consider again the estimates (5.14) and (5.15)

$$\begin{cases} \langle \mu_{\delta_n}, f \rangle - \dfrac{\alpha(\delta_n)}{2}\|\mu_{\delta_n}\|^2_{\mathcal{V}} \geqslant \langle \mu_0, f \rangle - \dfrac{\alpha(\delta_n)}{2}\|\mu_0\|^2_{\mathcal{V}}, \\ \langle \mu_0, f \rangle \geqslant \langle \mu_{\delta_n}, f \rangle \end{cases}$$

and let $n \to \infty$ so that $\delta_n \to 0$ and $\alpha(\delta_n) \to 0$. We get that

$$\begin{cases} \langle \mu^*, f \rangle \geqslant \langle \mu_0, f \rangle, \\ \langle \mu_0, f \rangle \geqslant \langle \mu^*, f \rangle. \end{cases}$$

Therefore, $\langle \mu^*, f\rangle = \langle \mu_0, f\rangle$ and μ^* solves the dual limit problem (5.12).

Using weak lower semicontinuity of the norm in a Hilbert space, from (5.16) we get that

$$\|\mu^*\|_{\mathcal{V}} \leqslant \liminf_{n\to\infty} \|\mu_{\delta_n}\|_{\mathcal{V}} \leqslant \|\mu_0\|_{\mathcal{V}} \tag{5.17}$$

for any μ_0 solving (5.12). Therefore, μ^* is the minimum norm solution of (5.12) (unique, since it is an orthogonal projection of zero onto the feasible set in (5.12)). From (5.17) and (5.16) with $\mu_0 = \mu^*$ we than also get that

$$\|\mu_{\delta_n}\|_{\mathcal{V}} \to \|\mu^*\|_{\mathcal{V}}$$

and the convergence $\mu_{\delta_n} \to \mu^*$ is actually strong by the Radon-Riesz property of the norm in a Hilbert space (see Remark 4.2.6). So, we get the assertion of the Theorem with $\mu^\dagger = \mu^*$. $\square$

Example 5.2.4 (Total Variation)**.** Let $\mathcal{U} = \mathcal{V} = L^2(\Omega)$ with $\Omega \subset \mathbb{R}^2$ bounded and $\mathcal{C} \subset \Omega$ a domain with a C^∞ boundary. Let $\mathcal{J}(\cdot) = \mathrm{TV}(\cdot)$ and $A\colon L^2(\Omega) \to L^2(\Omega)$ be the identity operator (i.e., we consider the problem of denoising). From Example 4.3.6 we know that

$$\mathrm{TV}(\mathbf{1}_{\mathcal{C}}) = \mathrm{Per}(\mathcal{C}),$$

where $\mathbf{1}_{\mathcal{C}}$ is the indicator function of the set $\mathcal{C}$. Denoting by $\mathbf{n}_{\partial\mathcal{C}}$ the unit normal, we obtain

$$\mathrm{Per}(\mathcal{C}) = \int_{\partial\mathcal{C}} 1 = \int_{\partial\mathcal{C}} \langle \mathbf{n}_{\partial\mathcal{C}}, \mathbf{n}_{\partial\mathcal{C}}\rangle .$$

Since $\mathbf{n}_{\partial\mathcal{C}} \in C^\infty(\partial\mathcal{C}, \mathbb{R}^2)$ and $\|\mathbf{n}_{\partial\mathcal{C}}(x)\|_2 = 1$ for any x, we can extend $\mathbf{n}_{\partial\mathcal{C}}$ to a $C_0^\infty(\Omega, \mathbb{R}^2)$ vector field ψ with $\sup_{x\in\Omega} \|\psi(x)\|_2 \leqslant 1$. Therefore, using the divergence theorem, we obtain that

$$\int_{\partial\mathcal{C}} \langle \mathbf{n}_{\partial\mathcal{C}}, \mathbf{n}_{\partial\mathcal{C}}\rangle = \int_{\partial\mathcal{C}} \langle \psi, \mathbf{n}_{\partial\mathcal{C}}\rangle = \int_{\mathcal{C}} \operatorname{div}\psi = \langle \operatorname{div}\psi, \mathbf{1}_{\mathcal{C}}\rangle .$$

Combining all these equalities, we get that

$$\mathrm{TV}(\mathbf{1}_{\mathcal{C}}) = \langle \operatorname{div}\psi, \mathbf{1}_{\mathcal{C}}\rangle .$$

Note that, since $\psi \in C_0^\infty(\Omega, \mathbb{R}^2)$, $\operatorname{div}\psi \in C^\infty(\Omega) \subset L^2(\Omega)$.

Taking an arbitrary $u \in \mathcal{U}$, we note that

$$\mathrm{TV}(u) - \langle \operatorname{div}\psi, u\rangle = \sup_{\substack{\varphi \in C_0^\infty(\Omega,\mathbb{R}^2) \\ \sup_{x\in\Omega}\|\varphi(x)\|_2 \leqslant 1}} \langle u, \operatorname{div}\varphi\rangle - \langle u, \operatorname{div}\psi\rangle \geqslant 0,$$

since $\varphi = \operatorname{div}\psi$ is feasible. Therefore, $\operatorname{div}\psi \in \partial\,\mathrm{TV}(0)$ and, since $\mathrm{TV}(\mathbf{1}_{\mathcal{C}}) = \langle \operatorname{div}\psi, \mathbf{1}_{\mathcal{C}}\rangle$, we also get that

$$\operatorname{div}\psi \in \partial\,\mathrm{TV}(\mathbf{1}_{\mathcal{C}}).$$

Since A is the identity operator, $\mathcal{R}(A^*) = \mathcal{U}$ and the source condition is satisfied at $u = \mathbf{1}_{\mathcal{C}}$ with $\mu = \operatorname{div}\psi$.

Example 5.2.5 (Total Variation)**.** Let $\mathcal{U} = \mathcal{V} = L^2(\Omega)$ with $\Omega \subset \mathbb{R}^2$ bounded and $\mathcal{C} \subset \Omega$ be a domain with a nonsmppth boundary, e.g., a square $\mathcal{C} = [0,1]^2$. Let $\mathcal{J}(\cdot) = \mathrm{TV}(\cdot)$. We will show in this example that in this case $\partial\,\mathrm{TV}(\mathbf{1}_{\mathcal{C}}) = \emptyset$.

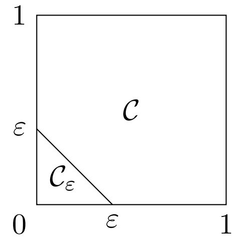

Figure 5.1: Example of a set whose indicator function does not satisfy the source condition.

Assume that there exists $p_0 \in \partial\,\mathrm{TV}(\mathbf{1}_\mathcal{C}) \subset L^2(\Omega)$. Then by the results of Example 4.3.6 we have that

$$\langle p_0, \mathbf{1}_\mathcal{C}\rangle = \mathrm{TV}(\mathbf{1}_\mathcal{C}) = \mathrm{Per}(\mathcal{C}) = 4.$$

Since p_0 is a subgradient, we get that for any $u \in L^2(\Omega)$

$$\mathrm{TV}(u) - \langle p_0, u\rangle \geqslant 0.$$

Let us cut a triangle $\mathcal{C}_\varepsilon$ of size ε from a corner of $\mathcal{C}$ as shown in Figure 5.1. Then for $u = \mathbf{1}_{\mathcal{C}\setminus\mathcal{C}_\varepsilon}$ we get

$$\mathrm{TV}(\mathbf{1}_{\mathcal{C}\setminus\mathcal{C}_\varepsilon}) \geqslant \left\langle p_0, \mathbf{1}_{\mathcal{C}\setminus\mathcal{C}_\varepsilon}\right\rangle = \langle p_0, \mathbf{1}_\mathcal{C}\rangle - \langle p_0, \mathbf{1}_{\mathcal{C}_\varepsilon}\rangle$$

and therefore

$$\langle p_0, \mathbf{1}_{\mathcal{C}_\varepsilon}\rangle \geqslant \mathrm{TV}(\mathbf{1}_\mathcal{C}) - \mathrm{TV}(\mathbf{1}_{\mathcal{C}\setminus\mathcal{C}_\varepsilon}) = \mathrm{Per}(\mathcal{C}) - \mathrm{Per}(\mathcal{C}\setminus\mathcal{C}_\varepsilon) = 4 - (4 - 2\varepsilon + \sqrt{2}\varepsilon) = (2-\sqrt{2})\varepsilon > 0.$$

By Hölder's inequality we get that

$$\langle p_0, \mathbf{1}_{\mathcal{C}_\varepsilon}\rangle = \int_{\mathcal{C}_\varepsilon} p_0 \cdot \mathbf{1} \leqslant \left(\int_{\mathcal{C}_\varepsilon} |p_0|^2\right)^{1/2} \left(\int_{\mathcal{C}_\varepsilon} 1\right)^{1/2} = \frac{1}{\sqrt{2}}\varepsilon \left(\int_{\mathcal{C}_\varepsilon} |p_0|^2\right)^{1/2}.$$

Combining the last two inequalities, we get

$$(2-\sqrt{2})\varepsilon \leqslant \langle p_0, \mathbf{1}_{\mathcal{C}_\varepsilon}\rangle \leqslant \frac{1}{\sqrt{2}}\varepsilon \left(\int_{\mathcal{C}_\varepsilon} |p_0|^2\right)^{1/2}$$

and therefore

$$\int_{\mathcal{C}_\varepsilon} |p_0|^2 \geqslant 2(2-\sqrt{2})^2 > 0$$

for all $\varepsilon > 0$. However, since $p_0 \in L^2(\Omega)$ by assumption, we must have

$$\int_{\mathcal{C}_\varepsilon} |p_0|^2 \to 0 \quad \text{as } \varepsilon \to 0.$$

This contradiction proves that such p_0 does not exist and $\partial\,\mathrm{TV}(\mathbf{1}_\mathcal{C}) = \emptyset$.

5.3 Convergence Rates

Now we are ready to answer the question that we asked in the beginning of this Chapter - *how fast* do the regularised solutions converge to a $\mathcal{J}$-minimising solution? The answer is given by the following Theorem.

Theorem 5.3.1. *Let the source condition* (5.13) *be satisfied at a $\mathcal{J}$-minimising solution $u^\dagger_{\mathcal{J}}$ and let u_δ be a regularised solution solving* (5.2)*. Then the following estimate holds*

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

In the particular case when $\alpha(\delta) \sim \delta$ we get

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

Proof. Consider the function

$$\varphi(g) = \frac{1}{2}\|g - f_\delta\|^2_{\mathcal{V}}.$$

It is differentiable and its subdifferential is given by

$$\partial\varphi(g) = \{g - f_\delta\}.$$

Taking $g = Au_\delta$ and using the definition of a Bregman distance from f to Au_δ we obtain

$$D^{Au_\delta - f_\delta}_{\varphi}(f, Au_\delta) = \frac{1}{2}\|f - f_\delta\|^2_{\mathcal{V}} - \frac{1}{2}\|Au_\delta - f_\delta\|^2_{\mathcal{V}} - \langle Au_\delta - f_\delta, f - Au_\delta\rangle \geqslant 0$$

and therefore

$$\langle Au_\delta - f_\delta, f - Au_\delta\rangle \leqslant \frac{1}{2}\|f - f_\delta\|^2_{\mathcal{V}} - \frac{1}{2}\|Au_\delta - f_\delta\|^2_{\mathcal{V}}.$$

Consider the symmetric Bregman distance $D^{symm}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}})$. We obtain the following expression

$$\begin{aligned}\alpha D^{symm}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) &= \alpha\left\langle A^*\mu^\dagger - A^*\mu_\delta, u^\dagger_{\mathcal{J}} - u_\delta\right\rangle = \alpha\left\langle \mu^\dagger - \mu_\delta, f - Au_\delta\right\rangle \\ &= \alpha\left\langle \mu^\dagger, f - Au_\delta\right\rangle + \langle -\alpha\mu_\delta, f - Au_\delta\rangle.\end{aligned}$$

From the optimality conditions (5.10) we know that $\alpha\mu_\delta = f_\delta - Au_\delta$. Therefore, we get

$$\begin{aligned}\alpha D^{symm}_{\mathcal{J}}(u_\delta, u^\dagger_{\mathcal{J}}) &= \alpha\left\langle \mu^\dagger, f - Au_\delta\right\rangle + \langle Au_\delta - f_\delta, f - Au_\delta\rangle \\ &\leqslant \alpha\left\langle \mu^\dagger, f - Au_\delta\right\rangle + \frac{1}{2}\|f - f_\delta\|^2_{\mathcal{V}} - \frac{1}{2}\|Au_\delta - f_\delta\|^2_{\mathcal{V}} \\ &\leqslant \alpha\left\langle \mu^\dagger, f_\delta - Au_\delta\right\rangle + \alpha\left\langle \mu^\dagger, f - f_\delta\right\rangle + \frac{1}{2}\delta^2 - \frac{1}{2}\|Au_\delta - f_\delta\|^2_{\mathcal{V}} \\ &\leqslant \alpha\|\mu^\dagger\|_{\mathcal{V}}\left(\|f_\delta - Au_\delta\|_{\mathcal{V}} + \|f - f_\delta\|_{\mathcal{V}}\right) + \frac{1}{2}\delta^2 - \frac{1}{2}\|Au_\delta - f_\delta\|^2_{\mathcal{V}} \\ &\leqslant \alpha\delta\|\mu^\dagger\|_{\mathcal{V}} - \left(\frac{1}{2}\|Au_\delta - f_\delta\|^2_{\mathcal{V}} - \alpha\|\mu^\dagger\|_{\mathcal{V}}\|Au_\delta - f_\delta\|_{\mathcal{V}} + \frac{1}{2}\alpha^2\|\mu^\dagger\|_{\mathcal{V}}\right) \\ &+ \frac{1}{2}\alpha^2\|\mu^\dagger\|_{\mathcal{V}} + \frac{1}{2}\delta^2 = \alpha\delta\|\mu^\dagger\|_{\mathcal{V}} - \frac{1}{2}\left(\|Au_\delta - f_\delta\|_{\mathcal{V}} - \alpha\|\mu^\dagger\|_{\mathcal{V}}\right)^2 \\ &+ \frac{1}{2}\alpha^2\|\mu^\dagger\|_{\mathcal{V}} + \frac{1}{2}\delta^2 \leqslant \frac{1}{2}\delta^2 + \alpha\delta\|\mu^\dagger\|_{\mathcal{V}} + \frac{1}{2}\alpha^2\|\mu^\dagger\|_{\mathcal{V}} = \frac{1}{2}\left(\delta + \alpha\|\mu^\dagger\|_{\mathcal{V}}\right)^2,\end{aligned}$$

which yields the desired estimate. With $\alpha(\delta) \sim \delta$ we immediately get

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

□

Chapter 6

Numerical Optimisation Methods

In the last two chapters we formulated several optimisation problems in the context of variational regularisation. In this Chapter we will discuss some methods that can be used to efficiently solve them numerically. We will consider the case when $\mathcal{U}$ is a Banach space and will study methods that can be used to find the minimum of a functional $E \colon \mathcal{U} \to \mathbb{R}$.

6.1 More on derivatives in Banach spaces

Let us first discuss some more properties of derivatives and subdifferentials.

Definition 6.1.1. *Let $E \colon \mathcal{U} \to \mathbb{R}$ be a mapping from the Banach space $\mathcal{U}$ to $\mathbb{R}$ and $u \in \mathcal{U}$. If there exists an operator $A \in \mathcal{L}(\mathcal{U}, \mathbb{R}) = \mathcal{U}^*$ that*

$$\lim_{h \to 0} \frac{|E(u+h) - E(u) - Ah|}{\|h\|_{\mathcal{U}}} = 0\,,$$

holds true, then E is called Fréchet differentiable *in u and $E'(u) := A$ the* Fréchet derivative *in u. If the Fréchet derivative exists for all $u \in \mathcal{U}$, the operator $E' : \mathcal{U} \to \mathcal{U}^*$ is called Fréchet differentiable.*

Example 6.1.2. Let $\mathcal{U}$ be a Banach space and $p \in \mathcal{U}^*$. Then the Fréchet derivative of $\langle u, p \rangle$ is $\langle u, p \rangle' = p$.

Example 6.1.3. Let $\mathcal{U}$ be a Hilbert space and $M \in \mathcal{L}(\mathcal{U}, \mathcal{U})$. Then the Fréchet derivative of $E \colon \mathcal{U} \to \mathbb{R}$,

$$E(u) = \|u\|_M^2 := \langle Mu, u \rangle$$

at any $u \in \mathcal{U}$ is given by

$$E'(u)(\cdot) = \langle (M + M^*)u, \cdot \rangle\ ,$$

and thus by the Riesz representation theorem can be identified with $(M + M^*)u$. If M is self-adjoint then $E'(u) = 2Mu$.

Proof. Simple calculations show that

$$E(u+h) - E(u) = \langle (M + M^*)u, h \rangle + \langle Mh, h \rangle$$

which shows that

$$\frac{|E(u+h) - E(u) - \langle (M + M^*)u, h \rangle|}{\|h\|_{\mathcal{U}}} = \frac{|\langle Mh, h \rangle|}{\|h\|_{\mathcal{U}}} \leqslant \frac{\|M\| \|h\|_{\mathcal{U}}^2}{\|h\|_{\mathcal{U}}} \to 0$$

for $\|h\| \to 0$. $\square$

Example 6.1.4. Let $\mathcal{U}, \mathcal{V}$ be Hilbert spaces, $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$, $f \in \mathcal{V}$ and $E \colon \mathcal{U} \to \mathbb{R}$ be defined as $E(u) := \frac{1}{2}\|Au - f\|_{\mathcal{V}}^2$. Then the Fréchet derivative of E can be identified with

$$E'(u) = A^*(Au - f)\,.$$

Proof. It is clear that

$$\frac{1}{2}\|Au - f\|_{\mathcal{V}}^2 = \frac{1}{2}\|Au\|_{\mathcal{V}}^2 - \langle Au, f\rangle + \frac{1}{2}\|f\|_{\mathcal{V}}^2 = \frac{1}{2}\langle u, A^*Au\rangle - \langle u, A^*f\rangle + \frac{1}{2}\|f\|_{\mathcal{V}}^2$$

and

$$E'(u) = A^*Au - A^*f = A^*(Au - f).$$

$\square$

An immediate consequence of Theorem 4.1.21 if the following

Proposition 6.1.5. *Let $\mathcal{U}$ be a normed space, $E \colon \mathcal{U} \to \mathbb{R}$ be convex and Fréchet differentiable and $F \colon \mathcal{U} \to \bar{\mathbb{R}}$ be proper, l.s.c. and convex. Then for all $u \in \operatorname{dom}(E+F) = \operatorname{dom}(F)$ it holds*

$$\partial(E + F)(u) = E'(u) + \partial F(u)\,.$$

6.2 Gradient descent

Let $\mathcal{U}$ be a Hilbert space. In this section we will analyse the iteration

$$u^{k+1} = u^k - \tau E'(u^k) \tag{6.1}$$

called *gradient descent*, which is one of the most popular methods to solve smooth minimisation problems.

Lemma 6.2.1 (Descent Lemma)**.** *Let $E \colon \mathcal{U} \to \mathbb{R}$ be Fréchet differentiable and E' Lipschitz continuous with constant $L \in \mathbb{R}$ (which we will call L-smooth in what is to follow). Then for all $x, y \in \mathcal{U}$ we have that*

$$E(x) \leqslant E(y) + \big\langle E'(y), x - y\big\rangle + \frac{L}{2}\|x - y\|^2\,.$$

Proof. For any $t \in [0, 1]$ define $g(t) := E(y + t(x - y))$ for which we obviously have $g(1) = E(x)$ and $g(0) = E(y)$. Then we have that

$$\begin{aligned}\int_0^1 \big\langle E'(y + t(x - y)) - E'(y), x - y\big\rangle\, dt &\leqslant \int_0^1 \|E'(y + t(x - y)) - E'(y)\|\|x - y\| dt \\ &\leqslant \int_0^1 Lt\|x - y\|^2 dt \\ &= \frac{L}{2}\|x - y\|^2\end{aligned}$$

and can further estimate

$$\begin{aligned}E(x) - E(y) = g(1) - g(0) &= \int_0^1 g'(t) dt \\ &= \int_0^1 \big\langle E'(y + t(x - y)), x - y\big\rangle\, dt \\ &= \int_0^1 \big\langle E'(y), x - y\big\rangle\, dt + \int_0^1 \big\langle E'(y + t(x - y)) - E'(y), x - y\big\rangle\, dt \\ &\leqslant \big\langle E'(y), x - y\big\rangle + \frac{L}{2}\|x - y\|^2\,.\end{aligned}$$

$\square$

Remark 6.2.2. If E is convex, then the inequality of the lemma can also be written in terms of the Bregman distance as $D_E^{E'(y)}(x, y) \leqslant \frac{L}{2}\|x-y\|^2$.

Theorem 6.2.3 (Convergence of gradient descent)**.** *Let E be L-smooth and the step size of gradient descent be chosen as*

$$\tau < \frac{2}{L}\,.$$

Then gradient descent monotonically decreases the function value, i.e.

$$E(u^{k+1}) \leqslant E(u^k)\,.$$

Moreover, if E is bounded from below, then the gradients convergence to zero, i.e.

$$\|E'(u^k)\| \to 0\,,$$

with rate (for some $C > 0$)

$$\min_{k=0,\dots,K-1} \|E'(u^k)\| \leqslant \frac{C}{K^{1/2}}\,.$$

Proof. Choosing $x = u^{k+1}$ and $y = u^k$ in the Descent Lemma yields

$$\begin{aligned} E(u^{k+1}) - E(u^k) &\leqslant \left\langle E'(u^k), -\tau E'(u^k) \right\rangle + \frac{L}{2}\|\tau E'(u^k)\|^2 \\ &= -\tau \|E'(u^k)\|^2 + \frac{\tau^2 L}{2}\|E'(u^k)\|^2 = -\frac{c}{2}\|E'(u^k)\|^2 \end{aligned} \tag{6.2}$$

with $c := \tau L\left(\frac{2}{L} - \tau\right) > 0$ which shows the monotonic descent.

Moreover, summing (6.2) over $k = 0, \dots, K-1$ yields

$$E(u^K) - E(u^0) \leqslant -\frac{c}{2}\sum_{k=0}^{K-1} \|E'(u^k)\|^2$$

and after rearranging

$$\sum_{k=0}^{K-1} \|E'(u^k)\|^2 \leqslant \frac{E(u^0) - E(u^K)}{c/2} \leqslant \frac{E(u^0) - \inf_{u \in \mathcal{U}} E(u)}{c/2} \leqslant C^2\,.$$

Thus, letting $K \to \infty$ we have that

$$\|E'(u^k)\| \to 0$$

and the convergence is with rate

$$\min_{k=0,\dots,K-1} \|E'(u^k)\|^2 \leqslant \frac{1}{K}\sum_{k=0}^{K-1} \|E'(u^k)\|^2 \leqslant \frac{C^2}{K}\,.$$

Taking the square root completes the proof. $\square$

Remark 6.2.4. It follows from the theorem that if $\{u^k\}_k$ converges, then it converges to a *stationary point* $u^* \in \mathcal{U}$ with $E'(u^*) = 0$.

Example 6.2.5. Consider Tikhonov regularisation, which consists in minimising

$$E(u) = \frac{1}{2}\|Au - f\|_{\mathcal{V}}^2 + \frac{\alpha}{2}\|u\|_{\mathcal{U}}^2$$

over all u in a Hilbert space $\mathcal{U}$ (where $\mathcal{V}$ is also a Hilbert space). Using the results of Examples 6.1.3 and 6.1.4, we get that

$$E'(u) = A^*(Au - f) + \alpha u.$$

We also observe that for all $u, v \in \mathcal{U}$

$$E'(u) - E'(v) = A^*(Au - f) + \alpha u - A^*(Av - f) - \alpha v = (A^*A + \alpha I)(u - v)$$

and

$$\|E'(u) - E'(v)\|_{\mathcal{V}} \leqslant \|A^*A + \alpha I\|_{\mathcal{L}(\mathcal{U},\mathcal{V})}\|u - v\|_{\mathcal{U}}.$$

Therefore, E' is Lipschitz continuous with constant $L = \|A^*A + \alpha I\|_{\mathcal{L}(\mathcal{U},\mathcal{V})}$ and we can use gradient descent with step size

$$\tau < \frac{2}{L}.$$

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(8):1258–1259, 1984.

[6] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2011.

[7] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111, 2018.

[8] B. Bollobás. *Linear Analysis: An Introductory Course*. Cambridge University Press, Cambridge, second edition, 1999.

[9] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411, 2004.

[10] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[11] N. Dunford and J. T. Schwartz. *Linear Operators, Part 1: General Theory*. Wiley Interscience Publishers, 1988.

[12] N. Dunford and J. T. Schwartz. *Linear Operators, Part 2: Spectral Theory, Self Adjoint Operators in Hilbert Space*. Wiley Interscience Publishers, 1988.

[13] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. 1976.

[14] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[15] C. W. Groetsch. *Stable approximate evaluation of unbounded operators*. Springer, 2006.

[16] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.

[17] J. Hunter and B. Nachtergaele. *Applied Analysis*. World Scientific Publishing Company Incorporated, 2001.

[18] J. A. Iglesias, G. Mercier, and O. Scherzer. A note on convergence of solutions of total variation regularized linear inverse problems. *Inverse Problems*, 34(5):055011, 2018.

[19] A. W. Naylor and G. R. Sell. *Linear Operator Theory in Engineering and Science*. Springer Science & Business Media, 2000.

[20] A. Rieder. *Keine Probleme mit Inversen Problemen: Eine Einführung in ihre stabile Lösung*. Vieweg+Teubner Verlag, 2003.

[21] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[22] W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991.

[23] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Springer, 2009.

[24] T. Tao. *Epsilon of Room, One*, volume 1. American Mathematical Soc., 2010.

[25] E. Zeidler. *Applied Functional Analysis: Applications to Mathematical Physics*, volume 108 of *Applied Mathematical Sciences Series*. Springer, 1995.

[26] E. Zeidler. *Applied Functional Analysis: Main Principles and Their Applications*, volume 109 of *Applied Mathematical Sciences Series*. Springer, 1995.