Inverse Problems

Martin Benning
Updated on: June 6, 2016

Lecture Notes
Michaelmas Term 2015

Contents

1 Introduction to inverse problems **5**
1.1 Examples . . . 5
1.1.1 Matrix inversion . . . 6
1.1.2 Differentiation . . . 7
1.1.3 Deconvolution . . . 8
1.1.4 Tomography . . . 8
1.1.5 Magnetic Resonance Imaging (MRI) . . . 10

2 Linear inverse problems **13**
2.1 Generalised inverse . . . 14
2.2 Compact operators . . . 17
2.3 Singular value decomposition of compact operators . . . 18

3 Regularisation **21**
3.1 Truncated singular value decomposition . . . 25
3.2 Tikhonov regularisation . . . 26
3.3 Source-conditions . . . 27
3.4 Asymptotic regularisation . . . 28
3.5 Landweber iteration . . . 29
3.6 Variational regularisation . . . 30
3.6.1 Tikhonov-type regularisation . . . 32
3.7 Convex Variational Calculus . . . 34

4 Computational realisation **39**
4.1 A primal dual hybrid gradient method . . . 39
4.1.1 Convergence of PDHGM . . . 41
4.1.2 Deconvolution with total variation regularisation . . . 43

These lecture notes are based on the following books and lecture notes:

1. Engl, Heinz Werner, Martin Hanke, and Andreas Neubauer. Regularization of inverse problems. Vol. 375. Springer Science & Business Media, 1996.

2. Mueller, Jennifer L., and Samuli Siltanen. Linear and nonlinear inverse problems with practical applications. Vol. 10. SIAM, 2012.

3. Martin Burger, Inverse Problems. Lecture notes winter 2007/2008.

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

4. Christian Clason, Inverse Probleme (in German), Lecture notes winter 2014/2015

   `https://www.uni-due.de/~adf040p/teaching/inverse_14/InverseSkript.pdf`

The lecture notes are under constant redevelopment and will likely contain errors and mistakes. I very much appreciate the finding and reporting of those (to mb941@cam.ac.uk). Thanks!

Chapter 1

Introduction to inverse problems

Solving an inverse problem is the task of computing an unknown physical quantity that is related to given, indirect measurements via a forward model. Inverse problems appear in a vast majority of applications, including imaging (Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Electron Tomography (ET), microscopic imaging, geophysical imaging), signal- and image-processing, computer vision, machine learning and (big) data analysis in general, and many more.
Mathematically, an inverse problem can be described as the solution of the operator equation

$$Ku = f \tag{1.1}$$

with given measurement data f for the unknown quantity u. Here, $K : \mathcal{U} \to \mathcal{V}$ denotes an operator mapping from the Banach space $\mathcal{U}$ to the Banach space $\mathcal{V}$. Throughout this lecture we are going to restrict ourselves to linear and bounded operators though.

Inverting a forward model however is not straightforward in most relevant applications, for two basic reasons: either a (unique) inverse model simply does not exist, or existing inverse models heavily amplify small measurement errors. In the sense of Hadamard the problem (1.1) is called *well-posed* if

- for all input data there exists a solution of the problem, i.e. for all $f \in \mathcal{V}$ there exists a $u \in \mathcal{U}$ with $Ku = f$.
- for all input data this solution is unique, i.e. $u \neq v$ implies $Kv \neq f$.
- the solution of the problem depends continuously on the input datum, i.e. for all $\{u_k\}_{k\in\mathbb{N}}$ with $Ku_k \to f$ we have $u_k \to u$.

If any of these conditions is violated, problem (1.1) is called *ill-posed*. In the following we are going to see that most practically relevant inverse problems are ill-posed or approximately ill-posed.[1]

1.1 Examples

In the following we are going to present various examples of inverse problems and highlight the challenges of dealing with them.

[1] In fact the name ill-posed problems may be a more suitable name for this lecture, as the real challenge is to deal with the ill-posedness of the inverse problems. However, the name inverse problems became more widely accepted for this field of mathematics.

1.1.1 Matrix inversion

One of the most simple (class of) inverse problems that arises from (numerical) linear algebra is the solution of linear systems. These can be written in the form of (1.1) with $u \in \mathbb{R}^n$ and $f \in \mathbb{R}^n$ being n-dimensional vectors with real entries and $K \in \mathbb{R}^{n\times n}$ being a matrix with real entries. We further assume K to be a symmetric, positive definite matrix. In that case we know from the spectral theory of symmetric matrices that there exist Eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and corresponding Eigenvectors $k_j \in \mathbb{R}^n$ for $j \in \{1, \ldots, n\}$ such that K can be written as

$$K = \sum_{j=1}^{n} \lambda_j k_j k_j^T \,. \tag{1.2}$$

Note that by dividing (1.1) on both sides of the equality with λ_n, (1.2) can be rescaled so that $\lambda_n = 1$ and $\lambda_1 = \kappa^{-1}$ hold. Here, κ denotes the so-called condition number $\kappa = \lambda_n/\lambda_1$ (in the case of $\lambda_1 \neq 0$) of the matrix K. It is well known from numerical linear algebra that the condition number is a measure of how stable (1.1) can be solved.

If we assume that we observe f^δ instead of f, with $\|f - f^\delta\|_2 \leq \delta$ (here $\|\cdot\|_2$ denotes the Euclidean norm), and further denote u^δ the solution of $Ku^\delta = f^\delta$, the difference between u^δ and the solution u of (1.1) reads as

$$u - u^\delta = \sum_{j=1}^{n} \lambda_j^{-1} k_j k_j^T \left(f - f^\delta\right) .$$

Therefore we can estimate

$$\left\|u - u^\delta\right\|_2^2 = \sum_{j=1}^{n} \lambda_j^{-2} \underbrace{\|k_j\|_2^2}_{=1} \left|k_j^T \left(f - f^\delta\right)\right|^2 \leq \lambda_1^{-2} \left\|f - f^\delta\right\|_2^2 ,$$

due to $\lambda_1 \leq \lambda_i$ for $i \neq 1$ and the orthogonality of the eigenvectors. Thus, taking the square root yields the estimate

$$\left\|u - u^\delta\right\|_2 \leq \kappa \left\|f - f^\delta\right\|_2 \leq \kappa\delta \,.$$

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 K. 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. Let us consider the matrix

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

and the vector $f = (1, 1)^T$. Then the solution of $Ku = f$ is simply given via $u = (1, 0)^T$. If we, however, consider the perturbed data $f^\delta = (99/100, 101/100)^T$ instead of f, the solution u^δ of $Ku^\delta = f^\delta$ is $u^\delta = (-19.01, 20)^T$.

1.1.2 Differentiation

Another classic inverse problem is differentiation. Assume we are given a function $f \in L^\infty([0,1])$ with $f(0) = 0$ for which we want to compute $u = f'$. These conditions are satisfied if and only if u and f satisfy the operator equation

$$f(y) = \int_0^y u(x)\,dx\,,$$

which can be written as the operator equation $Ku = f$ with the linear operator $(K\cdot)(y) := \int_0^y \cdot(x)\,dx$. As in the previous section, we assume that instead of f we observe a noisy version f^δ for which we further assume that the perturbation is additive, i.e. $f^\delta = f + n^\delta$ with $f \in C^1([0,1])$ and $n^\delta \in L^\infty([0,1])$. It is obvious that the derivative u only exists if the noise n^δ is differentiable. However, even in case n^δ is differentiable the error in the derivative can become arbitrarily large. Consider for instance the sequence $(\delta_j)_{j\in\mathbb{N}}$ with $\delta_j \to 0$ and the sequence of corresponding noise functions $n^{\delta_j} \in C^1([0,1]) \hookrightarrow L^\infty([0,1])$ with

$$n^{\delta_j}(x) := \delta_j \sin\left(\frac{kx}{\delta_j}\right)\,,$$

for a fixed but arbitrary number k. We on the one hand observe $\left\|n^{\delta_j}\right\|_{L^\infty([0,1])} = \delta_j \to 0$, but on the other hand have

$$u^{\delta_j}(x) = f'(x) + k\cos\left(\frac{kx}{\delta_j}\right)\,,$$

and therefore obtain the estimate

$$\left\|u - u^{\delta_j}\right\|_{L^\infty([0,1])} = \left\|\left(n^{\delta_j}\right)'\right\|_{L^\infty([0,1])} = k\,.$$

Thus, despite the noise in the data becoming arbitrarily small, the error in the derivative can become arbitrarily big (depending on k).

Note that considering a decreasing error in the norm of the Banach space $C^1([0,1])$ will yield a different result. If we have a sequence of noise functions with $\left\|n^{\delta_j}\right\|_{C^1([0,1])} = \delta^j \to 0$ instead, we can conclude

$$\left\|u - u^{\delta_j}\right\|_{L^\infty([0,1])} \le \left\|n^{\delta_j}\right\|_{C^1([0,1])} \to 0\,,$$

due to $C^1([0,1])$ being embedded in $L^\infty([0,1])$. In contrast to the previous example the sequence of functions $n^{\delta_j}(x) := \delta_j \sin(kx)$ for instance satisfies

$$\left\|n^{\delta_j}\right\|_{C^1([0,1])} = \sup_{x\in[0,1]}\left|n^{\delta_j}(x)\right| + \sup_{x\in[0,1]}\left|\left(n^{\delta_j}\right)'(x)\right| = (1+k)\delta_j \to 0\,.$$

However, for a fixed δ the bound on $\left\|u - u^\delta\right\|_{L^\infty([0,1])}$ can obviously still become fairly large compared to δ, depending on how large k is.

1.1.3 Deconvolution

An interesting problem that occurs in many imaging, image- and signal processing applications is the deblurring or *deconvolution* of signals from a known, linear degradation. Deconvolution of a signal can be modelled as solving the inverse problem of the convolution, which reads as

$$f(y) = (Ku)(y) := \int_{\mathbb{R}^n} u(x)g(y-x)\,dx\,. \tag{1.3}$$

Here f denotes the blurry image, u is the (unknown) true image and g is the function that models the degradation. Due to the Fourier convolution theorem we can rewrite (1.3) to

$$f = (2\pi)^{\frac{n}{2}}\mathcal{F}^{-1}\left(\mathcal{F}(u)\mathcal{F}(g)\right)\,. \tag{1.4}$$

with $\mathcal{F}$ denoting the Fourier transform

$$\mathcal{F}(u)(\xi) := (2\pi)^{-\frac{n}{2}}\int_{\mathbb{R}^n} u(x)\exp(-ix\cdot\xi)\,dx \tag{1.5}$$

and $\mathcal{F}^{-1}$ being the inverse Fourier transform

$$\mathcal{F}^{-1}(f)(x) := (2\pi)^{-\frac{n}{2}}\int_{\mathbb{R}^n} f(\xi)\exp(ix\cdot\xi)\,d\xi \tag{1.6}$$

It is important to note that the inverse Fourier transform is indeed the unique, inverse operator of the Fourier transform in the Hilbert-space L^2 due to the theorem of Plancherel. If we rearrange (1.4) to solve it for u we obtain

$$u = (2\pi)^{-\frac{n}{2}}\mathcal{F}^{-1}\left(\frac{\mathcal{F}(f)}{\mathcal{F}(g)}\right)\,, \tag{1.7}$$

and hence, we allegedly can recover u by simple division in the Fourier domain. However, we are quickly going to discover that this inverse problem is ill-posed and the division will lead to heavy amplifications of small errors.

Let u denote the image that satisfies (1.3). Further we assume that instead of the blurry image f we observe $f^\delta = f + n^\delta$ instead, and that u^δ is the solution of (1.7) with input datum f^δ. Hence, we observe

$$(2\pi)^{\frac{n}{2}}\left|u-u^\delta\right| = \left|\mathcal{F}^{-1}\left(\frac{\mathcal{F}(f-f^\delta)}{\mathcal{F}(g)}\right)\right| = \left|\mathcal{F}^{-1}\left(\frac{\mathcal{F}(n^\delta)}{\mathcal{F}(g)}\right)\right|\,. \tag{1.8}$$

As the convolution kernel g usually has compact support, $\mathcal{F}(g)$ will tend to zero for high frequencies. Hence, the denominator of (1.8) becomes fairly small, whereas the numerator will be non-zero as the noise is of high frequency. Thus, in the limit the solution will not depend continuously on the data and the convolution problem therefore be ill-posed.

1.1.4 Tomography

In almost any tomography application the underlying inverse problem is either the inversion of the Radon transform or of the X-ray transform in dimensions higher than two. For $u \in C_0^\infty(\mathbb{R}^n)$,

$s \in \mathbb{R}$ and $\theta \in S^{n-1}$, the Radon transform[2] $R : C_0^\infty(\mathbb{R}^n) \to C^\infty(S^{n-1} \times \mathbb{R})$ can be defined as the line integral operator

$$f(\theta, s) = (\mathcal{R}u)(\theta, s) = \int_{x \cdot \theta = s} u(x)\, dx \tag{1.9}$$

$$= \int_{\mathbb{R}} u(s\theta + t\theta^\perp)\, dt\,. \tag{1.10}$$

Here $\theta^\perp$ denotes a vector orthogonal to θ.

Example 1.2. Let $n = 2$. Then S^{n-1} is simply the unit sphere $S^1 = \{\theta \in \mathbb{R}^2 \mid \|\theta\|_2 = 1\}$. We can choose for instance $\theta = (\cos(\varphi), \sin(\varphi))^T$, $\varphi \in [0, 2\pi[$, and parametrise the Radon transform in terms of φ and s, i.e.

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

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 of that line to the centre of the coordinate system.

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 on a small distance Δt it is proportional to the intensity itself, the density and the distance, i.e.

$$\frac{I(s\theta + (t + \Delta t)\theta^\perp) - I(s\theta + t\theta^\perp)}{\Delta t} = -I(s\theta + t\theta^\perp)u(s\theta + t\theta^\perp)\,.$$

By taking the limit $\Delta t \to 0$ we end up with the ordinary differential equation

$$\frac{d}{dt} I(s\theta + t\theta^\perp) = -I(s\theta + t\theta^\perp)u(s\theta + t\theta^\perp)\,. \tag{1.11}$$

We now integrate (1.11) from $t = -R/2$, the position of the emitter, to $t = R/2$, the position of the detector, to obtain

$$\int_{-\frac{R}{2}}^{\frac{R}{2}} \frac{\frac{d}{dt} I(s\theta + t\theta^\perp)}{I(s\theta + t\theta^\perp)}\, dt = -\int_{-\frac{R}{2}}^{\frac{R}{2}} u(s\theta + t\theta^\perp)\, dt\,.$$

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

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

As we know the radiation intensity at both the emitter and the detector, we therefore know $f(\theta, s) := \log(I(s\theta - (R\theta^\perp)/2)) - \log(I(s\theta + (R\theta^\perp)/2))$ and we can write the estimation of the unknown density u as the inverse problem of the Radon transform (1.10) (if we further assume that u can be continuously extended to zero outside of the circle of radius R).

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

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 photons 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 (1.10) again. The line of integration is determined by the position of the detector pairs and the geometry of the scanner.

1.1.5 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging (MRI) is an imaging technique that allows to visualise the chemical composition of patients or materials. MRI scanners use strong magnetic fields and radio waves to excite subatomic particles (like protons) that subsequently emit radio frequency signals which can be measured by the radio frequency coils. In the following we want to briefly outline the mathematics of the acquisition process. Subsequently we are going to see that finding the unknown spin proton density basically leads to solving the inverse problem of the Fourier transform (1.5). The magnetisation of a so-called spin isochromat can be described by the Bloch equations[3]

$$\frac{d}{dt}\begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} = \begin{pmatrix} -\frac{1}{T_2} & \gamma B_z(t) & -\gamma B_y(t) \\ -\gamma B_z(t) & -\frac{1}{T_2} & \gamma B_x(t) \\ \gamma B_y(t) & -\gamma B_x(t) & -\frac{1}{T_1} \end{pmatrix} \begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{M_0}{T_1} \end{pmatrix}. \tag{1.12}$$

Here $M(t) = (M_x(t), M_y(t), M_z(t))$ is the nuclear magnetisation (of the spin isochromat), γ is the gyromagnetic ratio, $B(t) = (B_x(t), B_y(t), B_z(t))$ denotes the magnetic field experienced by the nuclei, T_1 is the longitudinal and T_2 the transverse relaxation time and M_0 the magnetisation in thermal equilibrium. If we define $M_{xy}(t) = M_x(t) + iM_y(t)$ and $B_{xy}(t) = B_x(t) + iB_y(t)$, we can rewrite (1.12) to

$$\frac{d}{dt}M_{xy}(t) = -i\gamma\left(M_{xy}(t)B_z(t) - M_z(t)B_{xy}(t)\right) - \frac{M_{xy}(t)}{T_2} \tag{1.13a}$$

$$\frac{d}{dt}M_z(t) = i\frac{\gamma}{2}\left(M_{xy}(t)\overline{B_{xy}}(t) - \overline{M_{xy}}(t)B_{xy}(t)\right) - \frac{M_z(t) - M_0}{T_1} \tag{1.13b}$$

with $\bar{\cdot}$ denoting the complex conjugate of $\cdot$.

If we assume for instance that $B = (0, 0, B_0)$ is just a constant magnetic field in z-direction, (1.13) reduces to the decoupled equations

$$\frac{d}{dt}M_{xy}(t) = -i\gamma B_0 M_{xy}(t) - \frac{M_{xy}(t)}{T_2}, \tag{1.14a}$$

$$\frac{d}{dt}M_z(t) = -\frac{M_z(t) - M_0}{T_1}. \tag{1.14b}$$

[3]Named after the Swiss born American physicist Felix Bloch (23 October 1905 - 10 September 1983)

It is easy to see that this system of equations (1.14) has the unique solution

$$M_{xy}(t) = e^{-t(i\omega_0 + 1/T_2)} M_{xy}(0) \tag{1.15a}$$

$$M_z(t) = M_z(0) e^{-\frac{t}{T_1}} + M_0 \left(1 - e^{-\frac{t}{T_1}}\right) \tag{1.15b}$$

for $\omega_0 := \gamma B_0$ denoting the Lamor frequency, and $M_{xy}(0)$, $M_z(0)$ being the initial magnetisations at time $t = 0$.

Rotating frame

Thus, for a constant magnetic background field in z-direction, B_0, M_{xy} basically rotates around the z-axis in clockwise direction with frequency ω_0 (if we ignore the T_2 decay for a moment). Rotating the x- and y-coordinate axes with the same frequency yields the representation of the Bloch equations in the so-called rotating frame. If we substitute $M^r_{xy}(t) := e^{i\omega_0 t} M_{xy}(t)$, $B^r_{xy}(t) := B_{xy}(t) e^{i\omega_0 t}$, $M^r_z(t) := M_z(t)$ and $B^r_z(t) := B_z(t)$, we obtain

$$\frac{d}{dt} M^r_{xy}(t) = -i\gamma \left(M^r_{xy}(t)(B^r_z(t) - B_0) - M^r_z(t) B^r_{xy}(t) \right) - \frac{M^r_{xy}(t)}{T_2} \tag{1.16a}$$

$$\frac{d}{dt} M^r_z(t) = i\frac{\gamma}{2} \left(M^r_{xy}(t) \overline{B^r_{xy}}(t) - \overline{M^r_{xy}}(t) B^r_{xy}(t) \right) - \frac{M^r_z(t) - M_0}{T_1} \tag{1.16b}$$

instead of (1.13).

Thus, if we assume the magnetic field to be constant with magnitude B_0 in z-direction within the rotating frame, i.e. $B^r(t) = (B^r_x(t), B^r_y(t), B_0)$, (1.16a) simplifies to

$$\frac{d}{dt} M^r_{xy}(t) = i\gamma M^r_z(t) B^r_{xy}(t) - \frac{M^r_{xy}(t)}{T_2} \,. \tag{1.17}$$

90° pulse

Now we assume that $B^r_x(t) = c$, c constant, and $B^r_y(t) = 0$ for $t \in [0, \tau]$, and $\tau \ll T_1$ and $\tau \ll T_2$. Then we can basically ignore the effect of $M^r_{xy}(t)/T_2$ and $(M^r_z(t) - M_0)/T_1$, and the Bloch equations in the rotating frame simplify to

$$\frac{d}{dt} \begin{pmatrix} M^r_x(t) \\ M^r_y(t) \\ M^r_z(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \omega \\ 0 & -\omega & 0 \end{pmatrix} \begin{pmatrix} M^r_x(t) \\ M^r_y(t) \\ M^r_z(t) \end{pmatrix} \tag{1.18}$$

with $\omega := \gamma c$, in matrix form with separate components. Assuming the initial magnetisations in the rotating frame to be zero in the x-y plane, i.e. $M^r_x(0) = 0$ and $M^r_y(0) = 0$, and constant in the z-plane with value $M^r_z(0)$, the solution of (1.18) can be written as

$$\begin{pmatrix} M^r_x(t) \\ M^r_y(t) \\ M^r_z(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega t) & \sin(\omega t) \\ 0 & -\sin(\omega t) & \cos(\omega t) \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ M^r_z(0) \end{pmatrix} . \tag{1.19}$$

Thus, equation (1.19) rotates the initial z-magnetisation around the x-axis by the angle $\theta := \omega t$. Note that if c and τ are chosen such that $\theta = \pi$, all magnetisation is rotated from the z-axis to the y-axis, i.e. $M^r_y(\tau) = M^r_z(0)$. In analogy, choosing $B_x(t) = 0$ and $B_y(t) = c$, all magnetisation can be shifted from the z- to the x-axis.

Signal acquisition

If the radio-frequency (RF) pulse is turned off and thus, $B_x^r(t) = 0$ and $B_y^r(t) = 0$ for $t > \tau$, the same coils that have been used to induce the RF pulse can be used to measure the x-y magnetisation. Since we measure a volume of the whole x-y net-magnetisation, the acquired signal equals

$$y(t) = \int_{\mathbb{R}^3} M(x,t)\,dx = \int_{\mathbb{R}^2} e^{-i\omega_0(x)t} M^r(x,t)\,dx \tag{1.20}$$

with $M(x,t)$ denoting $M_{xy}(t)$ for a specific spatial coordinate $x \in \mathbb{R}^3$ ($M^r(x,t)$ respectively). Using (1.15a) and assuming $\tau < t \ll T_2$, this yields

$$y(t) = \int_{\mathbb{R}^3} M_\tau(x) e^{-i\omega_0(x)t}\,dx\,, \tag{1.21}$$

with M_τ denoting the x-y-magnetisation at spatial location $x \in \mathbb{R}^3$ and time $t = \tau$. Note that $M_\tau = 0$ without any RF pulse applied in advance.

Signal recovery

The basic clue to allow for spatially resolving nuclear magnetic resonance spectrometry is to add a magnetic field $\hat{B}(t)$ to the constant magnetic field B_0 in z-direction that varies spatially over time. Then, (1.14a) changes to

$$\frac{d}{dt} M_{xy}(t) = -i\gamma(B_0 + \hat{B}(t)) M_{xy}(t) - \frac{M_{xy}(t)}{T_2}\,,$$

which, for initial value $M_{xy}(0)$, has the unique solution

$$M_{xy}(t) = e^{-i\gamma\left(B_0 t + \int_0^t \hat{B}(\tau)\,d\tau\right)} e^{-\frac{t}{T_2}} M_{xy}(0) \tag{1.22}$$

if we ensure $\hat{B}(0) = 0$. If now $x(t)$ denotes the spatial location of a considered spin isochromat at time t, we can write $\hat{B}(t)$ as $\hat{B}(t) = x(t) \cdot G(t)$, with a function G that describes the influence of the magnetic field gradient over time.
Based on the considerations that lead to (1.21) we therefore measure

$$y(t) = \int_{\mathbb{R}^3} M_\tau(x) e^{-i\gamma\left(B_0(x)t + \int_0^t x(\tau)\cdot G(\tau)\,d\tau\right)}\,dx$$

in an NMR experiment. Further assuming that B_0 is also constant in space, we can eliminate this term and write the signal acquisition as

$$e^{i\gamma B_0 t} y(t) = \int_{\mathbb{R}^3} M_\tau(x) e^{-i\gamma \int_0^t x(\tau)\cdot G(\tau)\,d\tau}\,dx \tag{1.23}$$

In the following we assume that $x(t)$ can be approximated reasonably well via its zero-order Taylor approximation around $t_0 = 0$, i.e.

$$\int_0^t x(\tau)\cdot G(\tau)\,d\tau \approx x(0)\cdot \int_0^t G(\tau)\,d\tau\,. \tag{1.24}$$

and hence, the inverse problem of finding the unknown spin-proton density M_τ for given measurements y is equivalent to solving the inverse problem of the Fourier transform

$$f(t) = (KM_\tau)(t) = \int_{\mathbb{R}^3} M_\tau(x) e^{-ix\cdot\xi(t)}\,dx\,, \tag{1.25}$$

with $f(t) := e^{i\gamma B_0 t} y(t)$ and $\xi(t) := \gamma \int_0^t G(\tau)\,d\tau$.

Chapter 2

Linear inverse problems

Throughout this lecture we deal with functional analytic operators. For the sake of brevity, we cannot recall all basic concepts of functional analysis but refer to popular textbooks that deal with this subject, like [10]. Nevertheless, we want to recall a few important properties that are going to be important for the further course of this lecture. In particular, we are going to focus on inverse problems with *bounded, linear operators* K only, i.e. $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ with

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

For $\mathcal{X} \subset \mathcal{U}$ and $K : \mathcal{X} \to \mathcal{V}$ we further want to denote with

1. $\mathcal{D}(K) := \mathcal{U}$ the domain
2. $\mathcal{N}(K) := \{u \in \mathcal{U} \mid Ku = 0\}$ the kernel
3. $\mathcal{R}(K) := \{Ku \in \mathcal{V} \mid x \in \mathcal{U}\}$ the range

of K. We say that K is continuous in $u \in \mathcal{U}$ if there exists a $\delta > 0$ for all $\varepsilon > 0$ with

$$\|Ku - Kv\|_{\mathcal{V}} \leq \varepsilon \text{ for all } v \in \mathcal{U} \text{ with } \|u - v\|_{\mathcal{U}} \leq \delta.$$

For linear K (as assumed throughout this lecture) it can be shown that continuity is equivalent to the existence of a positive constant C such that

$$\|Ku\|_{\mathcal{V}} \leq C\|u\|_{\mathcal{U}}$$

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

For the first part of the lecture we only consider $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ with $\mathcal{U}$ and $\mathcal{V}$ being Hilbert spaces. From functional calculus we know that every Hilbert space is equipped with a *scalar product*, which we are going to denote by $\langle \cdot, \cdot \rangle_{\mathcal{U}}$ (if $\mathcal{U}$ denotes the corresponding Hilbertspace). In analogy to the transpose of a matrix, this scalar product structure together with the theorem of Fréchet-Riesz [10, Section 2.10, Theorem 2.E] allows us to define the (unique) *adjoint operator* of K, denoted with K^*, as follows:

$$\langle Ku, v \rangle_{\mathcal{V}} = \langle u, K^* v \rangle_{\mathcal{U}} \,.$$

In addition to that, a scalar product allows to have a notion of orthogonality. Two function $u, v \in \mathcal{U}$ are said to be *orthogonal* if $\langle u, v\rangle_{\mathcal{U}} = 0$. For a subset $\mathcal{X} \subset \mathcal{U}$ the *orthogonal complement* of $\mathcal{X}$ in $\mathcal{U}$ is defined as

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

From this definition we immediately observe that $\mathcal{X}^{\perp}$ is a closed subspace. Further we have $\mathcal{U}^{\perp} = \{0\}$. Moreover, we have $\mathcal{X} \subset (\mathcal{X}^{\perp})^{\perp}$. If $\mathcal{X}$ is a closed subspace we even have $\mathcal{X} = (\mathcal{X}^{\perp})^{\perp}$ and $\{0\}^{\perp} = \mathcal{X}$. 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

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

see for instance [10, Section 2.9, Corollary 1]. The mapping $u \mapsto x$ defines a linear operator $P_{\mathcal{X}} \in \mathcal{L}(\mathcal{U}, \mathcal{X})$ that is called *orthogonal projection* on $\mathcal{X}$. The orthogonal projection satisfies the following conditions (cf. [6, Section 5.16]):

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

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

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

Hence, we can conclude the orthogonal decompositions

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

In the following we want to investigate the concept of generalised inverses of bounded linear operators, before we will identify compactness of operators as the major source of ill-posedness. Subsequently we are going to discuss this in more detail by analysing compact operators in terms of their singular value decomposition.

2.1 Generalised inverse

In order to overcome the issues of non-existence or non-uniqueness we want to generalise the concept of least-squares solutions to linear operators in Hilbert spaces.

If we consider the generic inverse problem (1.1) again, we know that there does not exist a solution of the inverse problem if $f \notin \mathcal{R}(K)$. In that case it seems reasonable to find an element $u \in \mathcal{U}$ for which $\|Ku - f\|_{\mathcal{V}}$ gets minimal instead.

However, for $\mathcal{N}(K) \neq \{0\}$ there are infinitely many solutions that minimise $\|Ku - f\|_{\mathcal{V}}$ of which we have to pick one. Picking the one with minimal norm $\|u\|_{\mathcal{U}}$ brings us to the definition of the minimal norm solution.

Definition 2.1. *An element $u \in \mathcal{U}$ is called*

- *least-squares solution of* (1.1)*, if*

$$\|Ku - f\|_{\mathcal{V}} = \min_{w \in \mathcal{U}} \|Kw - f\|_{\mathcal{V}} \,; \tag{2.1}$$

- *minimal norm solution of* (1.1)*, if*

$$\|u\|_{\mathcal{U}} = \min\{\|w\|_{\mathcal{U}} \mid w \text{ satisfies } (2.1)\} \,. \tag{2.2}$$

Note that for arbitrary f a minimum norm solution does not need to exist if $\mathcal{R}(K)$ is not closed. If, however, a minimum norm solution exists, it is unique and can (in theory) be computed via the *Moore-Penrose* generalised inverse.

Definition 2.2. *Let $\tilde{K} := K|_{\mathcal{N}(K)^\perp} : \mathcal{N}(K)^\perp \to \mathcal{R}(K)$ denote the restriction of $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then the Moore-Penrose inverse $K^\dagger$ is defined as the unique linear extension of $\tilde{K}^{-1}$ with*

$$\begin{aligned} \mathcal{D}(K^\dagger) &:= \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp \,, \\ \mathcal{N}(K^\dagger) &:= \mathcal{R}(K)^\perp \,. \end{aligned}$$

Note that $\tilde{K}$ is injective due to the restriction to $\mathcal{N}(K)^\perp$, and surjective due to the restriction to $\mathcal{R}(K)$. Hence, $\tilde{K}^{-1}$ exists, and - as a consequence - $K^\dagger$ is well-defined on $\mathcal{R}(K)$. Due to the orthogonal decomposition $\mathcal{D}(K^\dagger) = \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp$ there exist $f_1 \in \mathcal{R}(K)$ and $f_2 \in \mathcal{R}(K)^\perp$ with $f = f_1 + f_2$, for arbitrary $f \in \mathcal{D}(K^\dagger)$. Hence, we have

$$K^\dagger f = K^\dagger f_1 + K^\dagger f_2 = K^\dagger f_1 = \tilde{K}^{-1} f_1 \,, \tag{2.3}$$

due to $\mathcal{R}(K)^\perp = \mathcal{N}(K^\dagger)$, and thus, $K^\dagger$ is well-defined on the whole of $\mathcal{D}(K^\dagger)$. It can be shown that $K^\dagger$ can be characterized by the Moore-Penrose equations.

Lemma 2.1. *The Moore-Penrose inverse $K^\dagger$ satisfies $\mathcal{R}(K^\dagger) = \mathcal{N}(K)^\perp$ and the Moore-Penrose equations*

1. $KK^\dagger K = K$,
2. $K^\dagger K K^\dagger = K^\dagger$,
3. $K^\dagger K = I - P_{\mathcal{N}(K)}$,
4. $KK^\dagger = P_{\overline{\mathcal{R}(K)}}\big|_{\mathcal{D}(K^\dagger)}$,

where $P_{\mathcal{N}(K)} : \mathcal{U} \to \mathcal{N}(K)$ and $P_{\overline{\mathcal{R}(K)}} : \mathcal{V} \to \overline{\mathcal{R}(K)}$ denote the orthogonal projections on $\mathcal{N}(K)$ and $\overline{\mathcal{R}(K)}$, respectively.

Proof. First of all we are going to prove $\mathcal{R}(K^\dagger) = \mathcal{N}(K)^\perp$. According to (2.3) and the definition of the Moore-Penrose inverse we observe for $f \in \mathcal{D}(K^\dagger)$ that

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

as we not only have $P_{\overline{\mathcal{R}(K)}} f \in \overline{\mathcal{R}(K)}$ but also $P_{\overline{\mathcal{R}(K)}} f \in \mathcal{R}(K)$. Hence, $K^\dagger f \in \mathcal{R}(\tilde{K}^{-1}) = \mathcal{N}(K)^\perp$ and therefore $\mathcal{R}(K^\dagger) \subset \mathcal{N}(K)^\perp$. However, we also observe $\mathcal{N}(K)^\perp \subset \mathcal{R}(K^\dagger)$ as $K^\dagger K u = \tilde{K}^{-1}\tilde{K}u = u$ for $u \in \mathcal{N}(K)^\perp$ already implies $u \in \mathcal{R}(K^\dagger)$.

It remains to prove the Moore-Penrose equations. We begin with 4: For $f \in \mathcal{D}(K^\dagger)$ it follows from $\mathcal{R}(K^\dagger) = \mathcal{N}(K)^\perp$ and (2.4) that

$$K K^\dagger f = K\tilde{K}^{-1} P_{\overline{\mathcal{R}(K)}} f = \tilde{K}\tilde{K}^{-1} P_{\overline{\mathcal{R}(K)}} f = P_{\overline{\mathcal{R}(K)}} f \,,$$

due to $\tilde{K}^{-1} P_{\overline{\mathcal{R}(K)}} f \in \mathcal{N}(K)^\perp$ and $K = \tilde{K}$ on $\mathcal{N}(K)^\perp$.
3: According to the definition of $K^\dagger$ we have $K^\dagger K u = \tilde{K}^{-1} K u$ for all $u \in \mathcal{U}$ and thus

$$K^\dagger K u = \tilde{K}^{-1} \underbrace{K P_{\mathcal{N}(K)} u}_{=0} + \tilde{K}^{-1} K \underbrace{(I - P_{\mathcal{N}(K)})}_{=P_{\mathcal{N}(K)^\perp}} u = \tilde{K}^{-1}\tilde{K}(I - P_{\mathcal{N}(K)})u = (I - P_{\mathcal{N}(K)})u \,.$$

2: Inserting 4 into (2.4) yields

$$K^\dagger f = K^\dagger P_{\overline{\mathcal{R}(K)}} f = K^\dagger K K^\dagger f$$

for all $f \in \mathcal{D}(K^\dagger)$.
1: With 3 we have

$$K K^\dagger K = K(I - P_{\mathcal{N}(K)}) = K - K P_{\mathcal{N}(K)} = K \,.$$

□

The following theorem states that minimum norm solutions can be computed via the generalised inverse.

Theorem 2.1. *For each $f \in \mathcal{D}(K^\dagger)$ the minimal norm solution $u^\dagger$ of (1.1) is given via*

$$u^\dagger = K^\dagger f \,.$$

The set of all least squares solutions is given via $\{u^\dagger\} + \mathcal{N}(K)$.

Proof. See [2, Theorem 2.5]. □

In numerical linear algebra it is a well known fact that the normal equations can be considered to compute least squares solutions. The same is true in the continuous case.

Theorem 2.2. *For given $f \in \mathcal{D}(K^\dagger)$ the function $u \in \mathcal{U}$ is minimal norm solution if and only if it satisfies the normal equations*

$$K^* K u = K^* f \,. \tag{2.5}$$

We further have $u = u^\dagger$ in case of $u \in \mathcal{N}(K)^\perp$.

Proof. An element $u \in \mathcal{U}$ is least squares solution if and only if Ku is the projection of f onto $\mathcal{R}(K)$, i.e. $Ku = P_{\mathcal{R}(K)}(f)$. The latter is equivalent to $Ku \in \mathcal{R}(K)$ and $Ku - f \in \mathcal{R}(K)^\perp = \mathcal{N}(K^*)$. Thus, we conclude $K^*(Ku - f) = 0$. Further, a least squares solution has minimal norm if and only if $u \in \mathcal{N}(K)^\perp$. □

As a direct consequence from Theorem 2.1 and Theorem 2.2 we obtain

$$K^\dagger f = (K^*K)^\dagger K^* f\,,$$

and hence, in order to approximate $K^\dagger f$ we may also compute an approximation via (2.5) instead.

At the end of this section we further want to analyse the domain of the generalised inverse in more detail. Due to the construction of the Moore-Penrose inverse we have $\mathcal{D}(K^\dagger) = \mathcal{R}(K) \oplus \mathcal{R}(K)^\perp$. As orthogonal complements are always closed we can conclude

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

and hence, $\mathcal{D}(K^\dagger)$ is dense in $\mathcal{V}$. Thus, if $\mathcal{R}(K)$ is closed it follows that $\mathcal{D}(K^\dagger) = \mathcal{V}$ and vice versa, $\mathcal{D}(K^\dagger) = \mathcal{V}$ implies $\mathcal{R}(K)$ to be closed. Moreover, for $f \in \mathcal{R}(K)^\perp = \mathcal{N}(K^\dagger)$ the minimum-norm solution is $u^\dagger = 0$. Therefore, for given $f \in \overline{\mathcal{R}(K)}$ the important question to address is when f also satisfies $f \in \mathcal{R}(K)$. If this is the case, $K^\dagger$ has to be continuous. However, the existence of a single element $f \in \overline{\mathcal{R}(K)} \setminus \mathcal{R}(K)$ is enough already to prove that $K^\dagger$ is discontinuous.

Theorem 2.3. *Let $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then $K^\dagger \in \mathcal{L}(\mathcal{D}(K^\dagger), \mathcal{U})$ if and only if $\mathcal{R}(K)$ is closed.*

Proof. See [2, Proposition 2.4]. □

In the next section we are going to discover that the class of compact operators is a class for which the Moore-Penrose inverses are discontinuous.

2.2 Compact operators

Compact operators are very common in inverse problems; in fact, almost all (linear) inverse problems involve the inversion of compact operators. Compact operators are defined as follows.

Definition 2.3. *Let $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then K is said to be compact if the image of a bounded sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{U}$ contains a convergent subsequence $\{Ku_{n_k}\}_{k\in\mathbb{N}} \subset \mathcal{V}$. We denote the space of compact operators with $\mathcal{K}(\mathcal{U}, \mathcal{V})$.*

Compact operators can be seen as the infinite dimensional analogue to ill-conditioned matrices. Indeed it can be seen that compactness is a main source of ill-posedness, confirmed by the following result.

Theorem 2.4. *Let $K \in \mathcal{K}(\mathcal{U}, \mathcal{V})$ between infinite dimensional Banach space $\mathcal{U}$ and $\mathcal{V}$, such that the dimension of $\mathcal{R}(K)$ is infinite. Then the problem* (1.1) *is ill-posed, i.e. the corresponding Moore-Penrose inverse $K^\dagger$ is discontinuous.*

Proof. Since $\mathcal{U}$ and $\mathcal{R}(K)$ are infinite dimensional, we can conclude that $\mathcal{N}(K)^\perp$ is also infinite dimensional. We can therefore find a sequence $\{u_n\}_{n\in\mathbb{N}}$ with $u_n \in \mathcal{N}(K)^\perp$, $\|u_n\| = 1$ and $\langle u_n, u_k\rangle = 0$ for $n \neq k$. Since K is a compact operator the sequence $f_n = Ku_n$ is compact. Hence, we can find k, l such that for each $\delta > 0$ we have $\|Ku_l - Ku_k\| < \delta$. However, we also obtain

$$\|K^\dagger Ku_l - K^\dagger Ku_k\|^2 = \|u_l - u_k\|^2 = \|u_l\|^2 - \langle u_l, u_k\rangle + \|u_k\|^2 = 2\,.$$

Hence, $K^\dagger$ is discontinuous. □

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

2.3 Singular value decomposition of compact operators

We want to characterise the Moore-Penrose inverse of compact operators in terms of a spectral decomposition. Like in the finite dimensional case of matrices, we can only expect a spectral decomposition to exist for self-adjoint operators. Due to Theorem 2.2 we can consider K^*K instead of K, which brings us to the singular value decomposition of linear, compact operators.

Theorem 2.5. *Let $K \in \mathcal{K}(\mathcal{U}, \mathcal{V})$. Then there exists*

- *a null sequence $\{\sigma_j\}_{j\in\mathbb{N}}$ with $\sigma_1 \geq \sigma_2 \geq \ldots > 0$,*
- *an orthonormal basis $\{u_j\}_{j\in\mathbb{N}} \subset \mathcal{U}$ of $\overline{\mathcal{R}(K^*)}$,*
- *an orthonormal basis $\{v_j\}_{j\in\mathbb{N}} \subset \mathcal{V}$ of $\overline{\mathcal{R}(K)}$,*

with

$$Ku_j = \sigma_j v_j,\ K^*v_j = \sigma_j u_j \text{ for all } j \in \mathbb{N} \tag{2.6}$$

and

$$Kw = \sum_{j=1}^{\infty} \sigma_j \langle w, u_j\rangle_{\mathcal{U}}\, v_j \text{ for all } j \in \mathbb{N}. \tag{2.7}$$

A sequence $\{(\sigma_j, v_j, u_j)\}$ for which (2.7) is satisfied, is called singular system.

Proof. Exercise 3 on Exercise Sheet 2. □

Remark 2.1. Since Eigenvalues of K^*K with Eigenvectors u_j are also Eigenvalues of KK^* with Eigenvectors v_j, we further obtain a singular value decomposition of K^* due to (2.6), i.e.

$$K^*z = \sum_{j=1}^{\infty} \sigma_j \langle z, v_j\rangle_{\mathcal{V}}\, u_j \text{ for all } j \in \mathbb{N}.$$

We now want to derive a representation of the Moore-Penrose inverse in terms of the singular value decomposition. Remember that we have concluded $K^\dagger f = (K^*K)^\dagger K^* f$ from Theorem 2.1 and Theorem 2.2. Hence, for $u^\dagger = K^\dagger f$ we obtain

$$\sum_{j=1}^{\infty} \sigma_j^2 \langle u^\dagger, u_j\rangle_{\mathcal{U}}\, u_j = K^*Ku^\dagger = K^*f = \sum_{j=1}^{\infty} \sigma_j \langle f, v_j\rangle_{\mathcal{V}}\, u_j$$

and by comparison of the respective linear components

$$\langle u^\dagger, u_j\rangle_{\mathcal{U}} = \frac{1}{\sigma_j}\langle f, v_j\rangle_{\mathcal{V}}\,.$$

As a direct consequence, the singular value decomposition of the generalised inverse reads as

$$u^\dagger = K^\dagger f = \sum_{j=1}^{\infty} \frac{1}{\sigma_j}\langle f, v_j\rangle_{\mathcal{V}}\, u_j\,. \tag{2.8}$$

As for the singular value decomposition of K and K^* we have to verify if the sum converges. In contrast to the singular value decompositions of K and K^* this is not always true. Precisely, the so-called Picard criterion has to be met for (2.8) to converge.

Theorem 2.6. *Let $K \in \mathcal{K}(\mathcal{U}, \mathcal{V})$ with singular system $\{(\sigma_j, v_j, u_j)\}$, and $f \in \overline{\mathcal{R}(K)}$. Then $f \in \mathcal{R}(K)$ and (2.8) converges if and only if the Picard criterion*

$$\|K^\dagger f\|_{\mathcal{U}}^2 = \sum_{j=1}^{\infty} \frac{|\langle f, v_j \rangle_{\mathcal{V}}|^2}{\sigma_j^2} < \infty \tag{2.9}$$

is met.

Proof. See [2, Page 38, Theorem 2.8] □

Hence, the Picard criterion tells us how quickly the coefficients $\langle f, v_j \rangle_{\mathcal{V}}$ have to decay compared to the singular values σ_j. From representation (2.8) we can further conduct what happens in case of noisy measurements. Assume we have $f^\delta = f + \delta v_j$ instead of f. We then observe

$$\left\| K^\dagger f - K^\dagger f^\delta \right\|_{\mathcal{U}} = \delta \|K^\dagger v_j\|_{\mathcal{U}} = \frac{\delta}{\sigma_j} \to \infty \text{ for } j \to \infty \,.$$

For static j we see that the amplification of the error δ depends on how small σ_j is. Hence, the faster the singular values decay, the stronger the amplification of errors. For that reason, one distinguishes between three classes of ill-posed problems:

- Mildly ill-posed inverse problems are problems of the form (1.1) for which $c > 0$ and $r > 0$ exist with $\sigma_j \geq cj^{-r}$ for all $j \in \mathbb{N}$.
- Severely ill-posed inverse problems of the form (1.1) for which a significantly large $\tilde{j} \in \mathbb{N}$ exists, such that $\sigma_j \leq cj^{-r}$ is true for all $j \geq \tilde{j}$ and $c, r > 0$.
- Exponentially ill-posed problems of the form (1.1) for which a significantly large $\tilde{j} \in \mathbb{N}$ exists, such that $\sigma_j \leq ce^{-j^r}$ is true for all $j \geq \tilde{j}$ and $c, r > 0$.

Example 2.1. Let us consider the example of differentiation again, as introduced in Section 1.1.2. Let $\mathcal{U} = \mathcal{V} = L^2([0,1])$, then the operator K of the inverse problem (1.1) of differentiation is given as

$$(Ku)(y) = \int_0^y u(x)\,dx = \int_0^1 k(x,y)u(x)\,dx \,,$$

for

$$k(x,y) = \begin{cases} 1 & x \leq y \\ 0 & \text{else} \end{cases} \,.$$

From the theory of integral operators it follows that K is compact, due to it's kernel $k(x,y)$ being square integrable (cf. [1]). In order to compute the singular value decomposition we have to compute K^* first, which is characterised via

$$\langle Ku, v \rangle_{L^2([0,1])} = \langle u, K^* v \rangle_{L^2([0,1])} \,.$$

Hence, we obtain

$$\langle Ku, v \rangle_{L^2([0,1])} = \int_0^1 \int_0^1 k(x,y)u(x)\,dx\, v(y)\,dy = \int_0^1 u(x) \int_0^1 k(x,y)v(y)\,dy\,dx \,.$$

Hence, the adjoint operator K^* is given via

$$(K^*v)(x) = \int_0^1 k(x,y)v(y)\,dy = \int_x^1 v(y)\,dy\,.$$

Now we want to compute the Eigenvalues and Eigenvectors of K^*K, i.e. we look for $\lambda > 0$ and $u \in L^2([0,1])$ with

$$\lambda u(x) = (K^*Ku)(x) = \int_x^1 \int_0^y u(z)\,dz\,dy\,.$$

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

$$\lambda u'(x) = \frac{d}{dx}\int_x^1 \int_0^y u(z)\,dz\,dy = -\int_0^x u(z)\,dz\,,$$

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

$$\lambda u''(x) + u(x) = 0\,,$$

for which solutions are of the form

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

with $\sigma := \sqrt{\lambda}$ and constants c_1, c_2. In order to satisfy the boundary conditions $u(1) = 0$ and $u'(0) = 0$, the constants have to be chosen such that $c_1 = 0$ and $c_2 \cos(\sigma^{-1}) = 0$. Choosing $c_2 = 0$ would result in the trivial zero solution $u \equiv 0$, which is why we choose σ such that $\cos(\sigma^{-1}) = 0$ instead. 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 u_j:

$$u_j(x) = \sqrt{2}\cos\left(\left(j - \frac{1}{2}\right)\pi x\right)\,.$$

According to (2.6) we further obtain

$$v_j(x) = \sigma_j^{-1}(Ku_j)(x) = \left(j - \frac{1}{2}\right)\pi \int_0^x \sqrt{2}\cos\left(\left(j - \frac{1}{2}\right)\pi y\right)\,dy = \sqrt{2}\sin\left(\left(j - \frac{1}{2}\right)\pi x\right)\,,$$

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

$$\sum_{j=1}^{\infty} \frac{(2j-1)^2\pi^2}{2}\left|\int_0^1 f(x)\sin\left(\left(j - \frac{1}{2}\right)\pi x\right)\,dx\right|^2 < \infty\,.$$

Thus, the picard criterion in this case is equivalent to the condition that the Fourier series of f is differentiable with respect to every element of the series (and hence, that f is differentiable).

Chapter 3

Regularisation

We have seen in the previous section that the major source of ill-posedness of inverse problems of the type (1.1) is a fast decay of the singular values of K. An idea to overcome this issue is to define approximations of $K^\dagger$ in the following fashion. Consider the family of operators

$$R_\alpha f := \sum_{j=1}^{\infty} g_\alpha(\sigma_j)\langle f, v_j\rangle_{\mathcal{V}}\, u_j\,, \tag{3.1}$$

with functions $g_\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ that converge to $1/\sigma_j$ as α converges to zero. We are going to see that such an operator R_α is what is called a *regularisation* (of $K^\dagger$), if g_α is bounded, i.e.

$$g_\alpha(\sigma) \leq C_\alpha \text{ for all } \sigma \in \mathbb{R}_{>0}. \tag{3.2}$$

In case (3.2) holds true, we immediately observe

$$\|R_\alpha f\|_{\mathcal{U}}^2 = \sum_{j=1}^{\infty} g_\alpha(\sigma_j)^2 |\langle f, v_j\rangle_{\mathcal{V}}|^2 \leq C_\alpha^2 \sum_{j=1}^{\infty} |\langle f, v_j\rangle_{\mathcal{V}}|^2 \leq C_\alpha^2 \|f\|_{\mathcal{V}}^2\,,$$

which means that C_α is a bound for the norm of R_α.

Before we are going to examine concrete examples of regularisations of the form (3.1), we want to define what a regularisation actually is, and what properties come along with it.

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

- *$R_\alpha \in \mathcal{L}(\mathcal{V},\mathcal{U})$ for all $\alpha > 0$;*
- *$R_\alpha f \to K^\dagger f$ for all $f \in \mathcal{D}(K^\dagger)$.*

Hence, a regularisation is a pointwise approximation of the Moore-Penrose inverse with continuous operators. Note however, that we usually cannot expect $f \in \mathcal{D}(K^\dagger)$ for most applications, due to measurement and modelling errors. If we however assume that there exists $f \in \mathcal{D}(K^\dagger)$ such that we have

$$\left\| f - f^\delta \right\|_{\mathcal{V}} \leq \delta$$

for measured data $f^\delta \in \mathcal{V}$, we can expect that α has to be chosen in accordance to this error δ. Let us therefore consider the overall error $\|R_\alpha f^\delta - K^\dagger f\|_\mathcal{U}$ that we can split up as follows:

$$\begin{aligned}\left\|R_\alpha f^\delta - K^\dagger f\right\|_\mathcal{U} &\leq \left\|R_\alpha f^\delta - R_\alpha f\right\|_\mathcal{U} + \left\|R_\alpha f - K^\dagger f\right\|_\mathcal{U} \\ &\leq \delta \left\|R_\alpha\right\|_{\mathcal{L}(\mathcal{V},\mathcal{U})} + \left\|R_\alpha f - K^\dagger f\right\|_\mathcal{U} . \end{aligned} \tag{3.3}$$

The first term of (3.3) is the data error; this term unfortunately does not stay bounded for $\alpha \to 0$, which can be concluded from the uniform boundedness principle (also known as the Banach-Steinhaus theorem), cf. [5, Section 2.2, Theorem 2.2]. The second term however vanishes for $\alpha \to 0$, due to the pointwise convergence of R_α to $K^\dagger$. Hence it becomes evident from (3.3) that the choice of α depends on δ.

Definition 3.2. *A function $\alpha : \mathbb{R}_{>0} \times \mathcal{V} \to \mathbb{R}_{>0}$, $(\delta, f^\delta) \to \alpha(\delta, f^\delta)$ is called parameter choice rule. We distinguish between*

1. *a-priori parameter choice rules, if they depend on δ only;*
2. *a-posteriori parameter choice rules, if they depend on δ and f^δ;*
3. *heuristic parameter choice rules, if they depend on f^δ only.*

If $\{R_\alpha\}_{\alpha>0}$ is a regularisation of $K^\dagger$ and α is a parameter choice rule, then the pair (R_α, α) is called convergent regularisation, if for all $f \in \mathcal{D}(K^\dagger)$ there exists a parameter choice rule $\alpha : \mathbb{R}_{>0} \times \mathcal{V} \to \mathbb{R}_{>0}$ such that

$$\limsup_{\delta\to 0} \left\{ \left\|R_\alpha f^\delta - K^\dagger f\right\|_\mathcal{U} \;\middle|\; f^\delta \in \mathcal{V}, \left\|f - f^\delta\right\|_\mathcal{V} \leq \delta \right\} = 0 \tag{3.4}$$

and

$$\limsup_{\delta\to 0} \left\{ \alpha(\delta, f^\dagger) \;\middle|\; f^\delta \in \mathcal{V}, \left\|f - f^\delta\right\|_\mathcal{V} \leq \delta \right\} = 0$$

are guaranteed.

In case of 1 or 2 we would simply write $\alpha(\delta)$, respectively $\alpha(f^\delta)$, instead of $\alpha(\delta, f^\delta)$.

For the sake of brevity we are only discussing a-priori parameter choices throughout this lecture. In fact, it can be shown that for every regularisation an a-priori parameter choice rule, and thus, a convergent regularisation, exists.

Theorem 3.1. *Let $\{R_\alpha\}_{\alpha>0}$ be a regularisation of $K^\dagger$, for $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then there exists an a-priori parameter choice rule, such that (R_α, α) is a convergent regularisation.*

Proof. Let $f \in \mathcal{D}(K^\dagger)$ be arbitrary but fixed. We can find a monotone increasing function $\sigma : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that for every $\varepsilon > 0$ we have

$$\left\|R_{\sigma(\varepsilon)} f^\delta - K^\dagger f\right\|_\mathcal{U} \leq \frac{\varepsilon}{2} ,$$

due to the pointwise convergence $R_\alpha \to K^\dagger$.
As the operator $R_{\sigma(\varepsilon)}$ is continuous for fixed ε, there exists $\rho(\varepsilon) > 0$ with

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

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 $\mathcal{R}(\rho)$ 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

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

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

$$\left\|R_{\alpha(\delta)}f^\delta - K^\dagger f\right\|_{\mathcal{U}} \le \left\|R_{\sigma(\varepsilon)}f^\delta - R_{\sigma(\varepsilon)}f\right\|_{\mathcal{U}} + \left\|R_{\sigma(\varepsilon)}f^\delta - K^\dagger f\right\|_{\mathcal{U}} \le \varepsilon$$

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

Now we want to turn our attention to the case when $f \notin \mathcal{D}(K^\dagger)$. As the generalised inverse is not defined for those functions we cannot expect a convergent regularisation to stay bounded as $\alpha \to 0$. This is confirmed by the following result.

Theorem 3.2. *Let $K \in \mathcal{L}(\mathcal{U},\mathcal{V})$ and $\{R_\alpha\}_{\alpha>0}$ be a regularisation of $K^\dagger$. Then, $u_\alpha := R_\alpha f$ converges to $K^\dagger f$ as $\alpha \to 0$, for $f \in \mathcal{D}(K^\dagger)$. If*

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

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

Proof. The convergence in case of $f \in \mathcal{D}(K^\dagger)$ follows from Theorem 3.1. We therefore only need to consider the case $f \notin \mathcal{D}(K^\dagger)$. We assume that there exists a sequence $\alpha_k \to 0$ such that $\|u_{\alpha_k}\|_{\mathcal{U}}$ is uniformly bounded. Then there exists a weakly convergent subsequence $u_{\alpha_{k_l}}$ with some limit $u \in \mathcal{U}$, cf. [3, Section 2.2, Theorem 2.1]. As continuous linear operators are also weakly continuous, we further have $Ku_{\alpha_{k_l}} \rightharpoonup Ku$. However, as KR_α are uniformly bounded operators, we also conclude $Ku_{\alpha_{k_l}} = KR_{\alpha_{k_l}}f \rightharpoonup P_{\overline{\mathcal{R}(K)}}f$ (according to Lemma 2.1). Hence, we have $f \in \mathcal{D}(K^\dagger)$ in contradiction to the assumption $f \notin \mathcal{D}(K^\dagger)$. □

We can finally characterise a-priori parameter choice strategies that lead to convergent regularisation methods via the following theorem.

Theorem 3.3. *Let $\{R_\alpha\}_{\alpha>0}$ be a 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*

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

Proof. $\Rightarrow$: Let condition 1 and 2 be fulfilled. From (3.3) we then observe

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

Hence, (R_α, α) is a convergent regularisation method.
$\Leftarrow$: 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 1 is violated, $R_{\alpha(\delta)}f$ does not converge pointwise to $K^\dagger f$ and hence, (3.4) is violated for $f = f^\delta$ and $\delta = 0$. Hence, (R_α, α) is not a convergent regularisation method. If condition 1 is fulfilled but condition 2 is violated, there exists a null sequence $\{\delta_k\}_{k\in\mathbb{N}}$ with $\delta_k \|R_{\alpha(\delta_k)}\|_{\mathcal{L}(\mathcal{V},\mathcal{U})} \geq C > 0$, and hence, we can find a sequence $\{g_k\}_{k\in\mathbb{N}} \subset \mathcal{V}$ with $\|g_k\|_\mathcal{V} = 1$ and $\delta_k \|R_{\alpha(\delta_k)} g_k\|_\mathcal{U} \geq C$. Let $f \in \mathcal{D}(K^\dagger)$ be arbitrary and define $f_k := f + \delta_k g_k$. Then we have on the one hand $\|f - f_k\|_\mathcal{V} \leq \delta_k$, but on the other hand the norm of

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

cannot converge to zero, as the second term $\delta_k R_{\alpha(\delta_k)} g_k$ is bounded from below by construction. Hence, (3.4) is violated for $f^\delta = g_k$ and thus, (R_α, α) is not a convergent regularisation method. □

In the following we want to revisit (3.1) and prove that these methods are indeed regularisation methods for piecewise continuous functions g_α satisfying (3.2).

Theorem 3.4. *Let $g_\alpha : \mathbb{R}_{>0} \to \mathbb{R}$ be a piecewise continuous function satisfying* (3.2)*, $\lim_{\alpha\to 0} g_\alpha(\sigma) = \frac{1}{\sigma}$ and*

$$\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leq \gamma\,, \tag{3.5}$$

for some constant $\gamma > 0$. If R_α is defined as in (3.1)*, we have*

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

for all $f \in \mathcal{D}(K^\dagger)$.

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

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

From (3.5) we can conclude

$$\left| \left(\sigma_j g_\alpha(\sigma_j) - 1\right) \langle u^\dagger, u_j\rangle_\mathcal{U} \right| \leq (\gamma + 1) \|u^\dagger\|_\mathcal{U}\,,$$

and hence, each element of the sum stays bounded. Thus, we can estimate

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

and due to the pointwise convergence of $g_\alpha(\sigma_j)$ to $1/\sigma_j$ we deduce $\lim_{\alpha\to 0} \sigma_j g_\alpha(\sigma_j) - 1 = 0$. Hence, we have $\left\| R_\alpha f - K^\dagger f \right\|_\mathcal{U} \to 0$ for all $f \in \mathcal{D}(K^\dagger)$. □

Proposition 3.1. *Let the same assumptions hold as in Theorem 3.4. Further, let* (3.2) *be satisfied. Then (R_α, α) is a convergent regularisation method if*

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

is guaranteed.

Finally we want to consider the error propagation of a data error for the regularisation method (3.1).

Theorem 3.5. *Let the same assumptions hold for g_α as in Proposition 3.1. If we define $u_\alpha := R_\alpha f$ and $u_\alpha^\delta := R_\alpha f^\delta$, with $f \in \mathcal{D}(K^\dagger)$, $f^\delta \in \mathcal{V}$ and $\|f - f^\delta\|_\mathcal{V} \leq \delta$, then*

$$\|Ku_\alpha - Ku_\alpha^\delta\|_\mathcal{V} \leq \gamma\delta\,, \tag{3.6}$$

and

$$\|u_\alpha - u_\alpha^\delta\|_\mathcal{U} \leq C_\alpha\delta \tag{3.7}$$

hold true.

Proof. From the singular value decomposition we can estimate

$$\begin{aligned}\|Ku_\alpha - Ku_\alpha^\delta\|_\mathcal{V}^2 &\leq \sum_{j=1}^\infty \sigma_j^2 g_\alpha(\sigma_j)^2 |\langle f - f^\delta, v_j\rangle_\mathcal{V}|^2 \\ &\leq \gamma^2 \sum_{j=1}^\infty |\langle f - f^\delta, v_j\rangle_\mathcal{V}|^2 = \gamma^2 \|f - f^\delta\|_\mathcal{V}^2 \leq \gamma^2\delta^2\,,\end{aligned}$$

which yields (3.6). In the same fashion we can estimate

$$\begin{aligned}\|u_\alpha - u_\alpha^\delta\|_\mathcal{U}^2 &\leq \sum_{j=1}^\infty g_\alpha(\sigma_j)^2 |\langle f - f^\delta, v_j\rangle_\mathcal{V}|^2 \\ &\leq C_\alpha^2 \sum_{j=1}^\infty |\langle f - f^\delta, v_j\rangle_\mathcal{V}|^2 = C_\alpha^2 \|f - f^\delta\|_\mathcal{V}^2 \leq C_\alpha^2\delta^2\,,\end{aligned}$$

to obtain (3.7). □

Combining the assertions of Theorem 3.4, Proposition 3.1 and Theorem 3.5, we obtain the following convergence results of the regularised solutions.

Proposition 3.2. *Let the assumptions of Theorem 3.4, Proposition 3.1 and Theorem 3.5 hold true. Then,*

$$u_{\alpha(\delta, f^\delta)} \to u^\dagger$$

is guaranteed as $\delta \to 0$.

3.1 Truncated singular value decomposition

As a first example for a spectral regularisation of the form (3.1) we want to consider the so-called truncated singular value decomposition. As the name suggests, the idea is to discard all singular values below a certain threshold. If we identity this threshold with the regularisation parameter α, we can realise this truncation via

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

Note that we naturally obtain $\lim_{\alpha\to 0} g_\alpha(\sigma) = 1/\sigma$ for $\sigma > 0$. Equation (3.1) then reads as

$$u_\alpha = R_\alpha f = \sum_{\sigma_j \geq \alpha} \frac{1}{\sigma_j} \langle f, v_j \rangle_{\mathcal{V}} \, u_j \,, \tag{3.9}$$

for all $f \in \mathcal{V}$. Note that (3.9) is always finite for $\alpha > 0$ as zero is the only accumulation point of singular vectors of compact operators.

From (3.8) we immediately observe $g_\alpha(\sigma) \leq C_\alpha = 1/\alpha$. Thus, according to Proposition 3.1 the truncated singular value decomposition, together with an a-priori parameter choice strategy satisfying $\lim_{\delta\to 0} \alpha(\delta) = 0$, is a convergent regularisation method if $\lim_{\delta\to 0} \delta/\alpha(\delta) = 0$.

Moreover, we observe $\sup_{\sigma,\alpha} \sigma g_\alpha(\sigma) = \gamma = 1$ and hence, we obtain the error estimates $\|Ku_\alpha - Ku_\alpha^\delta\|_{\mathcal{V}} \leq \delta$ and $\|u_\alpha - u_\alpha^\delta\|_{\mathcal{U}} \leq \delta/\alpha$ as a consequence of Theorem 3.5.

3.2 Tikhonov regularisation

The main idea behind Tikhonov regularisation[1] is to shift the singular values of K^*K by a constant factor, which will be associated with the regularisation parameter α. This shift can be realised via the function

$$g_\alpha(\sigma) = \frac{\sigma}{\sigma^2 + \alpha} \,.$$

Again, we immediately observe $\lim_{\alpha\to 0} g_\alpha(\sigma) = 1/\sigma$ for $\sigma > 0$. Further, we can estimate $g_\alpha(\sigma) \leq 1/(2\sqrt{\alpha})$ due to $\sigma^2 + \alpha \geq 2\sqrt{\alpha}\sigma$. Moreover, we observe $\sigma g_\alpha(\sigma) = \sigma^2/(\sigma^2 + \alpha) < 1 =: \gamma$ for $\alpha > 0$. The corresponding Tikhonov regularisation (3.1) reads as

$$u_\alpha = R_\alpha f = \sum_{j=1}^{\infty} \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle f, v_j \rangle_{\mathcal{V}} \, u_j \,. \tag{3.10}$$

Note that Tikhonov regularisation can be computed without knowledge of the singular system. Considering the equation $(K^*K + \alpha I)u_\alpha$ in terms of the singular value decomposition, we observe

$$\begin{aligned} &\sum_{j=1}^{\infty} \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle f, v_j \rangle_{\mathcal{V}} \underbrace{K^* \underbrace{K u_j}_{=\sigma_j v_j}}_{=\sigma_j^2 u_j} + \sum_{j=1}^{\infty} \frac{\alpha \sigma_j}{\sigma_j^2 + \alpha} \langle f, v_j \rangle_{\mathcal{V}} \, u_j \\ &= \sum_{j=1}^{\infty} \frac{\sigma_j(\sigma_j^2 + \alpha)}{\sigma_j^2 + \alpha} \langle f, v_j \rangle_{\mathcal{V}} \, u_j = \sum_{j=1}^{\infty} \sigma_j \langle f, v_j \rangle_{\mathcal{V}} \, u_j = K^* f \,. \end{aligned}$$

Hence, the Tikhonov-regularised solution u_α can be obtained by solving

$$(K^*K + \alpha I)u_\alpha = K^* f \tag{3.11}$$

for u_α. The advantage in computing u_α via (3.11) is that its computation does not require the singular value decomposition of K, but only involves the inversion of a linear, well-posed operator equation with a symmetric, positive definite operator.

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

3.3 Source-conditions

Before we continue to investigate other examples of regularisation we want to briefly address the question of the convergence speed of a regularisation method. From Theorem 3.5 we have already obtained a convergence rate result; however, with additional regularity assumptions on the (unknown) minimal norm solution we are able to improve those. The regularity assumptions that we want to consider are known as *source conditions*, and are of the form

$$\exists w \in \mathcal{U} \ : \ u^\dagger = (K^*K)^\mu w \,. \tag{3.12}$$

The power $\mu > 0$ of the operator is understood in the sense of the consider the μ-th power of the singular values of the operator K^*K, i.e.

$$(K^*K)^\mu w = \sum_{j=1}^{\infty} \sigma_j^{2\mu} \langle w, u_j \rangle_{\mathcal{U}} u_j \,.$$

The rate of convergence of a regularisation scheme to the minimal norm solution now depends on the specific choice of g_α. We assume that g_α satisfies

$$\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leq \omega_\mu(\alpha) \,,$$

for all $\sigma > 0$. In case of the truncated singular value decomposition we would for instance have $\omega_\mu(\alpha) = \alpha^\mu$. With this additional assumption, we can improve the estimate in Theorem 3.4 as follows:

$$\begin{aligned} \|R_\alpha f - K^\dagger f\|_{\mathcal{V}} &\leq \sum_{j=1}^{\infty} |\sigma_j g_\alpha(\sigma_j) - 1|^2 |\langle u^\dagger, u_j \rangle_{\mathcal{U}}|^2 \\ &= \sum_{j=1}^{\infty} |\sigma_j g_\alpha(\sigma_j) - 1|^2 \sigma_j^{4\mu} |\langle w, u_j \rangle_{\mathcal{U}}|^2 \\ &\leq \omega_\mu(\alpha)^2 \|w\|_{\mathcal{U}}^2 \end{aligned}$$

Hence, we have obtained the estimate

$$\|u_\alpha - u^\dagger\|_{\mathcal{U}} \leq \omega_\mu(\alpha) \|w\|_{\mathcal{U}} \,.$$

Together with (3.3) we can further estimate

$$\|u_{\alpha(\delta)} - u^\dagger\|_{\mathcal{U}} \leq \omega_\mu(\alpha) \|w\|_{\mathcal{U}} + C_\alpha \delta \,. \tag{3.13}$$

Example 3.1. In case of the truncated singular value decomposition we know from Section 3.1 that $C_\alpha = 1/\alpha$, and we can further conclude $\omega_\mu(\alpha) = \alpha^{2\mu}$. Hence, (3.13) simplifies to

$$\|u_{\alpha(\delta)} - u^\dagger\|_{\mathcal{U}} \leq \alpha^{2\mu} \|w\|_{\mathcal{U}} + \delta \alpha^{-1} \tag{3.14}$$

in this case. In order to make the right-hand-side of (3.14) as small as possible, we have to choose α such that

$$\alpha = \left(\frac{\delta}{2\mu \|w\|_{\mathcal{U}}} \right)^{\frac{1}{2\mu+1}} \,.$$

With this choice of α we estimate

$$\begin{aligned}\|u_{\alpha(\delta)} - u^\dagger\|_{\mathcal{U}} &\leq \underbrace{2^{\frac{1-2\mu}{1+2\mu}}}_{\leq 2} \underbrace{\mu^{\frac{1-2\mu}{1+2\mu}}}_{\leq 1} \delta^{\frac{2\mu}{2\mu+1}} \|w\|_{\mathcal{U}}^{\frac{1}{2\mu+1}} \\ &\leq 2\delta^{\frac{2\mu}{2\mu+1}} \|w\|_{\mathcal{U}}^{\frac{1}{2\mu+1}} .\end{aligned}$$

It is important to note that no matter how large μ is, the rate of convergence $\delta^{\frac{2\mu}{2\mu+1}}$ will always be slower than δ, due to the ill-posedness of the inversion of K.

3.4 Asymptotic regularisation

Another form of regularisation is asymptotic regularisation of the form

$$\begin{aligned}\partial_t u(t) &= K^* (f - Ku(t)) \\ u(0) &= 0\end{aligned} . \tag{3.15}$$

As the linear operator K does not change with respect to the time t, we can make the Ansatz of writing $u(t)$ in terms of the singular value decomposition of K as

$$u(t) = \sum_{j=1}^{\infty} \gamma_j(t) u_j , \tag{3.16}$$

for some function $\gamma : \mathbb{R} \to \mathbb{R}$. From the initial conditions we immediately observe $\gamma(0) = 0$. From the singular value decomposition and (3.15) we further see

$$\sum_{j=1}^{\infty} \gamma_j'(t) u_j = \sum_{j=1}^{\infty} \sigma_j \left(\langle f, v_j \rangle_{\mathcal{V}} - \sigma_j \gamma(t) \underbrace{\langle u_j, u_j \rangle_{\mathcal{U}}}_{=\|u_j\|_{\mathcal{U}}^2 = 1} \right) u_j .$$

Hence, by equating the coefficients we get

$$\gamma_j'(t) = \sigma_j \langle f, v_j \rangle_{\mathcal{V}} - \sigma_j^2 \gamma_j(t) ,$$

and together with $\gamma_j(0)$ we obtain

$$\gamma_j(t) = \left(1 - e^{-\sigma_j^2 t}\right) \frac{1}{\sigma_j} \langle f, v_j \rangle_{\mathcal{V}}$$

as a solution for all j and hence, (3.16) reads as

$$u(t) = \sum_{j=1}^{\infty} \left(1 - e^{-\sigma_j^2 t}\right) \frac{1}{\sigma_j} \langle f, v_j \rangle_{\mathcal{V}} u_j .$$

If we substitute $t = 1/\alpha$, we obtain the regularisation

$$u_\alpha = \sum_{j=1}^{\infty} \left(1 - e^{-\frac{\sigma_j^2}{\alpha}}\right) \frac{1}{\sigma_j} \langle f, v_j \rangle_{\mathcal{V}} u_j$$

with $g_\alpha(\sigma) = \left(1 - e^{-\frac{\sigma^2}{\alpha}}\right) \frac{1}{\sigma}$. We immediately see that $g_\alpha(\sigma)\sigma \leq 1 =: \gamma$, and due to $e^x \geq 1 + x$ we further observe $1 - e^{-\frac{\sigma^2}{\alpha}} \leq \sigma^2/\alpha$ and therefore $(1 - e^{-\frac{\sigma^2}{\alpha}})/\sigma \leq \max_j \sigma_j/\alpha = \sigma_1/\alpha = \|K\|_{\mathcal{L}(\mathcal{U},\mathcal{V})}/\alpha =: C_\alpha$.

3.5 Landweber iteration

If we approximate (3.15) via a forward finite-difference discretisation, we end up with the iterative procedure

$$\begin{aligned} \frac{u^{k+1}-u^k}{\tau} &= K^*\left(f-Ku^k\right), \qquad (3.17)\\ \Leftrightarrow \quad u^{k+1} &= u^k + \tau K^*\left(f-Ku^k\right),\\ \Leftrightarrow \quad u^{k+1} &= (I-\tau K^*K)u^k + \tau K^* f, \end{aligned}$$

for some $\tau > 0$ and $u^0 \equiv 0$. Iteration (3.17) is known as the so-called Landweber iteration. We assume $f \in \mathcal{D}(K^\dagger)$ first, and with the singular value decomposition of K and K^* we obtain

$$\sum_{j=1}^{\infty}\langle u^{k+1}, u_j\rangle_{\mathcal{U}} u_j = \sum_{j=1}^{\infty}\left(\left(1-\tau\sigma_j^2\right)\langle u^k, u_j\rangle_{\mathcal{U}} + \tau\sigma_j\langle f, v_j\rangle_{\mathcal{V}}\right)u_j\,, \qquad (3.18)$$

and hence, by equating the individual summands

$$\langle u^{k+1}, u_j\rangle_{\mathcal{U}} = \left(1-\tau\sigma_j^2\right)\langle u^k, u_j\rangle_{\mathcal{U}} + \tau\sigma_j\langle f, v_j\rangle_{\mathcal{V}}\,. \qquad (3.19)$$

Assuming $u^0 \equiv 0$, summing up equation (3.19) yields

$$\langle u^k, u_j\rangle_{\mathcal{U}} = \tau\sigma_j\langle f, v_j\rangle_{\mathcal{V}}\sum_{i=1}^{k}(1-\tau\sigma_j^2)^{k-i}\,. \qquad (3.20)$$

The following Lemma will help us simplifying (3.20).

Lemma 3.1. *For $k \in \mathbb{N}\setminus\{1\}$ we have*

$$\sum_{i=1}^{k}(1-\tau\sigma^2)^{k-i} = \frac{1-\left(1-\tau\sigma^2\right)^k}{\tau\sigma^2}\,. \qquad (3.21)$$

Proof. Equation (3.21) can simply be verified via induction. We immediately see that

$$\sum_{i=1}^{2}(1-\tau\sigma^2)^{2-i} = 1+(1-\tau\sigma^2) = \frac{1-(1-2\tau\sigma^2+\tau^2\sigma^4)}{\tau\sigma^2} = \frac{1-\left(1-\tau\sigma^2\right)^2}{\tau\sigma^2}$$

serves as as our induction base. Considering $k \to k+1$, we observe

$$\begin{aligned} \sum_{i=1}^{k+1}(1-\tau\sigma^2)^{k+1-i} &= 1+\sum_{i=1}^{k}(1-\tau\sigma^2)^{k+1-i}\\ &= 1+(1-\tau\sigma^2)\sum_{i=1}^{k}(1-\tau\sigma^2)^{k-i}\\ &= 1+(1-\tau\sigma^2)\frac{1-\left(1-\tau\sigma^2\right)^k}{\tau\sigma^2}\\ &= \frac{1-\left(1-\tau\sigma^2\right)^{k+1}}{\tau\sigma^2}\,, \end{aligned}$$

and we are done. □

If we now insert (3.21) into (3.20) we therefore obtain

$$\langle u^k, u_j\rangle_{\mathcal{U}} = \left(1 - (1-\tau\sigma_j^2)^k\right) \frac{1}{\sigma_j}\langle f, v_j\rangle_{\mathcal{V}}. \tag{3.22}$$

The important consequence of Equation (3.22) is that we now immediately see that $\langle u^k, u_j\rangle_{\mathcal{U}} \to \langle u^\dagger, u_j\rangle_{\mathcal{U}}$ if we ensure $(1-\tau\sigma_j^2)^k \to 0$. In other words, we need to choose τ such that $|1-\tau\sigma_j^2| < 1$ (respectively $0 < \tau\sigma_j < 2$) for all j. As in the case of asymptotic regularisation we exploit that $\sigma_1 = \|K\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} > \sigma_j$ for all j and select τ such that

$$0 < \tau < \frac{2}{\|K\|^2_{\mathcal{L}(\mathcal{U},\mathcal{V})}}$$

is satisfied. If we interpret the iteration number as the regularisation parameter $\alpha := 1/k$, we obtain the regularisation method

$$u_\alpha = R_\alpha f = \sum_{j=1}^{\infty} \left(1 - \left(1-\tau\sigma_j^2\right)^{\frac{1}{\alpha}}\right) \frac{1}{\sigma_j}\langle f, v_j\rangle_{\mathcal{V}}$$

with $g_\alpha(\sigma) = \left(1 - (1-\tau\sigma^2)^{\frac{1}{\alpha}}\right)/\sigma$.

3.6 Variational regularisation

Variational regularisation aims at finding approximations to (1.1) by minimising appropriate functionals of the form

$$E_\alpha(u) := H(Ku, f) + \alpha J(u)\,. \tag{3.23}$$

Here $H : \mathcal{V} \times \mathcal{V} \to \mathbb{R} \cup \{+\infty\}$ and $J : \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$ represent two functionals over Banach spaces $\mathcal{U}$ and $\mathcal{V}$, $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ a linear and continuous operator, and $\alpha > 0$ is a real, positive constant. The term H is usually named *fidelity* or *data term*, as it measures the deviation between the measured data f and the forward model Ku. The functional J is the *regularisation term* as it will impose certain regularity conditions on the unknown u. The *regularisation parameter* will balance between both terms. It is beyond the scope of this lecture to discuss all the functional analytic concepts to study the minimisation of (3.23) for general functionals H and J. However, we briefly want to discuss three properties that will guarantee existence of a minimiser of (3.23). The first of them is a rather trivial one, assuring that functionals have some effective domain.

Definition 3.3. *A functional $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is called proper, if the effective domain*

$$\mathrm{dom}(E) := \{x \in \mathcal{X} \mid E(x) < \infty\}$$

is not empty.

More importantly we are going to need that functionals are lower semi-continuous. Roughly speaking this means that the functional values for arguments near an argument x are either close to $E(x)$ or greater than $E(x)$.

Definition 3.4. *Let $\mathcal{X}$ be a Banach space with topology τ. The functional $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is said to be lower semi-continuous at $x \in \mathcal{X}$ if*

$$E(x) \le \liminf_{k\to\infty} E(x_k)$$

for all $x_k \to x$ in the topology τ of $\mathcal{X}$.

Together with compactness of the sub level sets this leads to the fundamental theorem of optimisation.

Theorem 3.6. *Let $\mathcal{X}$ be a Banach space with topology τ and let $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous. Furthermore, let the level set*

$$\{x \in \mathcal{X} \mid E(x) \leq c\}$$

be non-empty, and compact in the topology τ for some $c \in \mathbb{R}$. Then there exists a global minimiser $\hat{x}$ of E, i.e. $E(\hat{x}) \leq E(x)$ for all $x \in \mathcal{X}$.

Proof. Let $\hat{E} = \inf_{x \in \mathcal{X}} E(x)$. Then a sequence $\{x_k\}_{k \in \mathbb{N}}$ exists with $E(x_k) \to \hat{E}$ for $k \to \infty$. For k sufficiently large, $E(x_k) \leq c$ holds and hence, $\{x_k\}_{k \in \mathbb{N}}$ is contained in a compact set. As a consequence, a subsequence $\{x_{k_l}\}_{l \in \mathbb{N}}$ exists with $x_{k_l} \to \hat{x}$, for $l \to \infty$, for some $\hat{x} \in \mathcal{X}$. From the lower semi-continuity of E we obtain

$$\hat{E} \leq E(\hat{x}) \leq \liminf_{l \to \infty} E(x_{k_l}) \leq \hat{E}\,,$$

consequently $\hat{x}$ is a global minimizer. □

For the sake of brevity, we will not going to prove lower semi-continuity and compactness of the sub-level sets for any of the functionals that we are going to consider. Nevertheless we want to emphasise (without proof) that all functionals considered throughout this lecture are lower semi-continuous and have compact sub level sets in the weak-* topology of the corresponding Banach space. According to the Theorem of Banach-Alaoglu (cf. [9, Theorem 1.9.13]), the set $\{v \in \mathcal{X}^* \mid \|v\|_{\mathcal{X}^*} \leq c\}$, for $c > 0$, is compact in the weak-* topology. Hence, we could conclude existence of a global minimum for a given infinite dimensional optimisation problem, if we were able to prove lower semi-continuity in the weak-* topology. Unfortunately this is not as easy as it sounds and certainly beyond the scope of this lecture.

Before we want to proceed to our first example of a variational regularisation model, we further want to recall the concepts of directional and Fréchet-derivatives of functionals.

Definition 3.5. *Let $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be a functional or operator. The directional derivative (also called first variation) at position $x \in \mathcal{X}$ in direction $y \in \mathcal{X}$ is defined as*

$$d_y E(x) := \lim_{t \downarrow 0} \frac{E(x + ty) - E(x)}{t}\,, \tag{3.24}$$

if that limit exists.

Note that if we define the function $\varphi_y(\tau) := E(x + \tau y)$, the directional derivative of E is equivalent to $\varphi_y'(\tau)|_{\tau=0}$

Definition 3.6. *Let $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ a functional mapping from the Banach space $\mathcal{X}$, and let $d_y E(x)$ exist for all $y \in \mathcal{X}$. If $E'(x) : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ exists such that*

$$(E'(x))(y) = d_y E(x) \quad \forall y \in \mathcal{X}\,,$$

and

$$\frac{|E(x + y) - E(x) - (E'(x))(y)|}{\|y\|_{\mathcal{X}}} \to 0 \quad \textit{for} \quad \|y\|_{\mathcal{X}} \to 0\,,$$

holds true, E is called Fréchet-differentiable in x and $E'(x)$ the Fréchet-derivative in x. If the Fréchet-derivative exists for all $x \in \mathcal{X}$, the operator $E' : \mathcal{X} \to \mathcal{X}^*$ is called Fréchet-derivative.

3.6.1 Tikhonov-type regularisation

First of all we want to show that the solution of the Tikhonov regularisation (3.11) can be interpreted as the minimiser of a functional of the form (3.23).

Theorem 3.7. *For $f \in \mathcal{V}$ the Tikhonov-regularised solution $u_\alpha = R_\alpha f$ with R_α as defined in (3.10) is uniquely determined as the global minimiser of the Tikhonov-functional*

$$T_\alpha(u) := \frac{1}{2}\|Ku - f\|_{\mathcal{V}}^2 + \frac{\alpha}{2}\|u\|_{\mathcal{U}}^2 . \tag{3.25}$$

Proof. $\Rightarrow$: A global minimiser $\hat{u} \in \mathcal{U}$ of $T_\alpha(\hat{u})$ is characterised via $T_\alpha(\hat{u}) \leq T_\alpha(u)$ for all $u \in \mathcal{U}$. Hence, it follows from

$$\begin{aligned} T_\alpha(u) - T_\alpha(u_\alpha) &= \frac{1}{2}\|Ku - f\|_{\mathcal{V}}^2 + \frac{\alpha}{2}\|u\|_{\mathcal{U}}^2 - \frac{1}{2}\|Ku_\alpha - f\|_{\mathcal{V}}^2 - \frac{\alpha}{2}\|u_\alpha\|_{\mathcal{U}}^2 \\ &= \frac{1}{2}\|Ku\|_{\mathcal{V}}^2 - \langle Ku, f\rangle + \frac{\alpha}{2}\|u\|_{\mathcal{U}}^2 - \frac{1}{2}\|Ku_\alpha\|_{\mathcal{V}}^2 + \langle Ku_\alpha, f\rangle - \frac{\alpha}{2}\|u_\alpha\|_{\mathcal{U}}^2 \\ &\quad + \underbrace{\langle (K^*K + \alpha I)u_\alpha - K^*f, u_\alpha - u\rangle}_{=0} \\ &= \frac{1}{2}\|Ku - Ku_\alpha\|_{\mathcal{V}}^2 + \frac{\alpha}{2}\|u - u_\alpha\|_{\mathcal{U}}^2 \\ &\geq 0 \end{aligned}$$

that u_α is a global minimiser of T_α.
$\Leftarrow$: If we have $T_\alpha(\hat{u}) \leq T_\alpha(u)$ (for all $u \in \mathcal{U}$), it follows with $u = \hat{u} + \tau v$ for arbitrary $\tau > 0$ and fixed $v \in \mathcal{U}$ that

$$0 \leq T_\alpha(u) - T_\alpha(\hat{u}) = \frac{\tau^2}{2}\|Kv\|_{\mathcal{V}}^2 + \frac{\tau^2\alpha}{2}\|v\|_{\mathcal{U}}^2 + \tau\langle (K^*K + \alpha I)\hat{u} - K^*f, v\rangle_{\mathcal{U}}$$

holds true. Dividing by τ and subsequent consideration of the limit $\tau \downarrow 0$ thus yields

$$\langle (K^*K + \alpha I)\hat{u} - K^*f, v\rangle_{\mathcal{U}} \geq 0\,, \text{ for all } v \in \mathcal{U}.$$

Hence, we can conclude $\hat{u} = u_\alpha$ and that u_α is the unique minimiser of (3.25). □

We therefore see, that u_α satisfying (3.10) (resp. (3.11)) is also the solution of a variational regularisation problem of the form (3.23), with $H(Ku, f) = \frac{1}{2}\|Ku - f\|_{\mathcal{V}}^2$ and $J(u) = \frac{1}{2}\|u\|_{\mathcal{U}}$. A nice property of this representation as a minimiser of a functional is that we can generalise Tikhonov regularisation by choosing different regularisation functionals for J, i.e.

$$\tilde{T}_\alpha(u) := \frac{1}{2}\|Ku - f\|_{\mathcal{V}}^2 + \alpha J(u)\,, \tag{3.26}$$

for general $J : \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$. We want to denote this regularisation as *Tikhonov-type regularisation* and present a few examples in the following. Note that Tikhonov-type regularisation for general J is not necessarily a regularisation in the sense of Definition 3.1, as we cannot expect $u_\alpha := \arg\min_{u\in\mathcal{U}} \tilde{T}_\alpha(u) \to u^\dagger$ for $\alpha \to 0$. However, we can generalise Definition 2.1 to justify calling the minimisation of (3.26) a regularisation.

Definition 3.7. *An element $u \in \mathcal{U}$ is called J-minimising solution of (1.1), if*

$$J(u) = \min\{J(w) \mid w \text{ satisfies } (2.1)\}\,. \tag{3.27}$$

Clearly, for f^δ with $\|f - f^\delta\|_{\mathcal{V}} \leq \delta$ and $u^\dagger$ satisfying (3.27) we can expect $u_\alpha \to u^\dagger$ for $\alpha \to 0$. However, it is beyond the scope of this lecture to verify this or any other generalised concepts of regularisation and we leave it to the interested reader to explore more about possible generalisations of regularisation.

Before we are going to restrict ourselves to convex functionals and address key properties of them and their minimisation, we want to give a few examples of regularisation functionals used in the context of Tikhonov-type regularisation.

Example 3.2 (Squared norm with linear operator)**.** The easiest way to extend classical Tikhonov regularisation to a more general regularisation method is to replace $\|u\|_{\mathcal{U}}^2/2$ with

$$J(u) := \frac{1}{2}\|Du\|_{\mathcal{W}}^2 ,$$

for an operator $D \in \mathcal{L}(\mathcal{U}, \mathcal{W})$. An example for an operator D that is commonly used in this framework is the gradient operator $\nabla \in \mathcal{L}(L^2(\Omega), L^2(\Omega; \mathbb{R}^n))$. In this context it is also common to use the H^1-norm as a regulariser, i.e.

$$J(u) = \frac{1}{2}\|u\|_{H^1(\Omega)} := \frac{1}{2}\|u\|_{L^2(\Omega)}^2 + \frac{1}{2}\|\nabla u\|_{L^2(\Omega;\mathbb{R}^n)}^2 .$$

Example 3.3 (Maximum-entropy regularisation)**.** Maximum-entropy regularisation is of particular interest if solutions of the inverse problem are assumed to be probability density functions, i.e. functions in the space

$$\mathrm{PDF}(\Omega) := \left\{ u \in L^1(\Omega) \;\middle|\; \int_\Omega u(x)\,dx = 1,\ u \geq 0 \right\} .$$

The (negative) entropy used in physics and information theory is defined as the functional $J : \mathrm{PDF}(\Omega) \to \mathbb{R}_{\geq -1}$ with

$$J(u) := \int_\Omega u(x)\log(u(x)) - u(x)\,dx ,$$

and the convention $0\log(0) := 0$. The corresponding Tikhonov-type regularisation reads as

$$u_\alpha = \operatorname*{arg\,min}_{u \in \mathrm{PDF}(\Omega)} \left\{ \frac{1}{2}\|Ku - f\|_{\mathcal{V}}^2 + \alpha \int_\Omega u(x)\log(u(x)) - u(x)\,dx \right\} , \tag{3.28}$$

for operators $K \in \mathcal{L}(\mathrm{PDF}(\Omega), \mathcal{V})$.

Example 3.4 (ℓ^1 regularisation)**.** When it comes to non-injective operators $K \in \mathcal{L}(\ell^2, \ell^2)$ between sequence spaces, the ℓ^1-norm

$$J(u) = \|u\|_{\ell^1} := \sum_{j=1}^\infty |u_j|$$

is often used as a regulariser, in order to enforce sparse solutions. The corresponding Tikhonov-type regularisation reads as

$$u_\alpha \in \operatorname*{arg\,min}_{u \in \ell^1} \left\{ \frac{1}{2}\|Ku - f\|_{\ell^2}^2 + \alpha \sum_{j=1}^\infty |u_j| \right\} .$$

Note that ℓ^p spaces are increasing in p, which implies $\|u\|_{\ell^2} \leq \|u\|_{\ell^1}$ and hence, $u \in \ell^2$ if $u \in \ell^1$.

Example 3.5 (Total variation regularisation)**.** Total variation as a regulariser has originally been introduced for image-denoising and -restoration applications with the goal to preserve edges in images, respectively discontinuities in signals [8]. For smooth signals $u \in W^{1,1}(\Omega)$ the total variation is simply defined as

$$J(u) := \int_\Omega \|(\nabla u)(x)\|_2 \, dx \,;$$

a more rigorous definition to also include functions with discontinuties is given via the dual formulation

$$J(u) = \mathrm{TV}(u) := \sup_{\substack{\varphi \in C_0^\infty(\Omega;\mathbb{R}^n) \\ \| \|\varphi\|_2 \|_\infty \leq 1}} \int_\Omega u(x) \, (\mathrm{div}\varphi)(x) \, dx \,.$$

Note that the choice of the vector norm is somewhat arbitrary and could be replaced by other p-vector norms. The corresponding Tikhonov-type regularisation reads as

$$u_\alpha \in \operatorname*{arg\,min}_{u \in \mathrm{BV}(\Omega)} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{V}}^2 + \alpha \mathrm{TV}(u) \right\} , \tag{3.29}$$

for $K \in \mathcal{L}(\mathrm{BV}(\Omega), \mathcal{V})$, and where $\mathrm{BV}(\Omega)$ denotes the space of *functions of bounded variation* defined as

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

3.7 Convex Variational Calculus

Definition 3.8 (Convex Set)**.** *Let $\mathcal{X}$ be a Banach space. A subset $\mathcal{C} \subseteq \mathcal{X}$ is called convex, if*

$$\lambda x + (1-\lambda) y \in \mathcal{C} \,,$$

for all $\lambda \in [0,1]$ and all $x, y \in \mathcal{C}$.

In analogy we can define (strictly) convex functionals on convex sets.

Definition 3.9 (Convex Functional)**.** *Let $\mathcal{C}$ be a convex set. A functional $E : \mathcal{C} \to \mathbb{R} \cup \{+\infty\}$ is called convex, if*

$$E(\lambda x + (1-\lambda) y) \leq \lambda E(x) + (1-\lambda) E(y) \tag{3.30}$$

for all $\lambda \in [0,1]$ and all $x, y \in \mathcal{C}$. The functional E is called strictly convex, if the equality of (3.30) only holds for $x = y$ or $\lambda \in \{0,1\}$.

Example 3.6 (Absolute Value Function)**.** For $\mathcal{C} = \mathbb{R}$ and $E : \mathbb{R} \to \mathbb{R}_{\geq 0}$ the absolute value function $E(x) = |x|$ is convex, since the absolute value function is a metric, and the triangular inequality yields $|\lambda x + (1-\lambda) y| \leq \lambda |x| + (1-\lambda)|y|$. Obviously, the absolute value function is not strictly convex.

Example 3.7 (Characteristic Function)**.** The characteristic function $\chi_{\mathcal{C}} : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ with

$$\chi_{\mathcal{C}}(x) := \begin{cases} 0 & x \in \mathcal{C} \\ +\infty & x \notin \mathcal{C} \end{cases} \tag{3.31}$$

is convex if $\mathcal{C} \subseteq \mathcal{X}$ is a convex subset.

We want to generalise the notion of differentiability to non-differentiable but convex functionals.

Definition 3.10 (Subdifferential)**.** *Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$, and let the functional $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be convex. Then, E is called subdifferentiable at $x \in \mathcal{X}$, if there exists an element $p \in \mathcal{X}^*$ such that*

$$E(y) - E(x) - \langle p, y - x \rangle_{\mathcal{X}} \geq 0$$

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

$$\partial E(x) := \{p \in \mathcal{X}^* \mid E(y) - E(x) - \langle p, v - u \rangle_{\mathcal{X}} \geq 0\,, \forall y \in \mathcal{X}\} \subset \mathcal{X}^*\,,$$

is called subdifferential of E at x.

For non-differentiable functionals the subdifferential is multivalued; we want to consider the subdifferential of the absolute value function as an illustrative example.

Example 3.8. Let $\mathcal{X} = \mathbb{R}$, and let $E : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be the absolute value function $E(u) = |x|$. Then, the subdifferential of E at x is given by

$$\partial E(x) = \operatorname{sign}(x) := \begin{cases} \{1\} & \text{for } x > 0 \\ [-1, 1] & \text{for } x = 0 \\ \{-1\} & \text{for } x < 0 \end{cases}.$$

This can easily be verified by case differentiation. For $x = 0$ the inequality $|x| \geq px$ is obviously fulfilled for any $x \in \mathbb{R}$ iff $p \in [-1, 1]$. For $x > 0$ the inequality $|y| \geq py + (1-p)x$ is fulfilled for every $y \in \mathbb{R}$ iff $p = 1$. In analogy, $p = -1$ has to hold for $x < 0$ in order to guarantee $|y| \geq py - (1+p)x$ for each $y \in \mathbb{R}$.

Throughout the remainder of this lecture we are particularly interested in convex, non-differentiable and absolutely one-homogeneous functionals (like the ℓ^1-norm or the total variation), i.e. we want to consider functionals E for which $E(cx) = |c|E(x)$ is satisfied, for every $c \in \mathbb{R}$. We therefore characterise the subdifferential of (absolutely) one-homogeneous functionals with the following lemma.

Lemma 3.2 (Subdifferential for One-Homogeneous Functionals)**.** *Let $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be a convex and one-homogeneous functional. Then, the subdifferential of E at x can equivalently be written as*

$$\partial E(x) = \{p \in \mathcal{X}^* \mid \langle p, x \rangle = E(x), \langle p, y \rangle_{\mathcal{X}} \leq E(y), \forall y \in \mathcal{X}\}\,.$$

Proof. If we consider $y = 0 \in \mathcal{X}$ in the definition of the subdifferential we immediately see

$$\langle p, x \rangle \geq E(x)\,,$$

while for $y = 2x \in \mathcal{X}$ we obtain

$$\langle p, x \rangle \leq E(2x) - E(x) = 2E(x) - E(x) = E(x)\,,$$

due to the one-homogeneity of E. Thus, $\langle p, x \rangle = E(x)$ follows. Moreover, if we insert $\langle p, x \rangle = E(x)$ in the definition of the subdifferential we get

$$\langle p, y \rangle \leq E(y)\,,$$

for all $y \in \mathcal{X}$. □

Example 3.9. If we consider the one-homogeneous absolute value function $E : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $E(x) = |x|$ again, we see that $xp = |x|$ implies $p = \operatorname{sign}(x)$. Moreover, we discover that for arguments y with the same sign as x we obtain $yp = |y|$, while for arguments y having the opposite sign as x, we get $yp = -|y| < |y|$.

Considering (3.23) again, we now and for the rest of the lecture assume H and J to be convex (in addition to the assumption of being proper and lower semi-continuous). Further, we assume H to be Fréchet-differentiable with respect to u. For this setup we can state the following useful characterisation of the subdifferential (without proof).

Theorem 3.8. *Let $H : \mathcal{V} \times \mathcal{V} \to \mathbb{R} \cup \{+\infty\}$ and $J : \mathcal{U} \to \mathbb{R} \cup \{+\infty\}$ be proper, convex and lower semi-continuous functionals, for Banach spaces $\mathcal{U}$ and $\mathcal{V}$. Let furthermore $K \in \mathcal{L}(\mathcal{U}, \mathcal{V})$, and H be Fréchet-differentiable with respect to u. Then, for $\alpha > 0$, the subdifferential of the functional E_α as defined in* (3.23) *is given via*

$$\partial(H(K\cdot, f) + \alpha J)(u) = K^* H'(Ku, f) + \alpha \partial J(u)\,.$$

Proof. See [7] or [4, Proposition 5.6]. □

Finally we want to characterise global minimisers of convex functionals.

Theorem 3.9 (Generalised Fermat)**.** *Let $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex functional. An element $x \in \mathcal{X}$ is a (global) minimiser of E iff $0 \in \partial E(x)$.*

Proof. For $0 \in \partial E(x)$ we have

$$0 = \langle 0, y - x\rangle \leq E(y) - E(x)$$

for all $y \in \mathcal{X}$ by definition of the subdifferential. Thus, x is a global minimizer of E.
If $0 \notin \partial E(x)$ holds, then there exists at least one $y \in \mathcal{X}$ such that

$$E(y) - E(x) < \langle 0, y - x\rangle = 0\,,$$

and hence, x cannot be a minimizer of E. □

It is worth mentioning that all the definitions and theorems derived in this chapter can be transferred to (strictly) concave functions and (global) maximisation problems. In particular, if E is a (strictly) concave functional then $-E$ is (strictly) convex, and the characterisation of a global maximum of E is identical to the characterisation of a global minimum of $-E$. We will make use of this fact throughout the next chapter when it comes to saddle-point problems.

To conclude this chapter, we want to introduce the useful notion of the generalised Bregman distance.

Definition 3.11. *For a proper, lower semi-continuous and convex functional $E : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ with non-empty subdifferential ∂E, the generalised Bregman distance is defined as*

$$D_E^p(x, y) := E(x) - E(y) - \langle p, x - y\rangle\,,\ p \in \partial E(y)\,. \tag{3.32}$$

The generalised Bregman distance is no distance in the classical sense. It does satisfy $D_E^p(x, y) \geq 0$ for all $x, y \in \mathcal{X}$; however, it is neither symmetric nor does it satisfy a triangular inequality in general. Nevertheless can we symmetrise the Bregman distance by simply adding the two Bregman distances with interchanged arguments.

Definition 3.12. *Let the same assumptions hold as in Definition 3.11. The generalised symmetric Bregman distance is defined as*

$$D_E^{symm}(x_1, x_2) := D_E^{p_1}(x_2, x_1) + D_E^{p_2}(x_1, x_2) = \langle x_1 - x_2, p_1 - p_2 \rangle ,$$

for $p_1 \in \partial E(x_1)$ *and* $p_2 \in \partial E(x_2)$.

Note that we have omitted p_1 and p_2 in the notation of D_E^{symm} only for the sake of brevity.

The generalised symmetric Bregman distance will be a helpful tool throughout the next chapter for the convergence analysis of the so-called primal-dual hybrid gradient method for the iterative solution of convex variational regularisation problems.

Chapter 4

Computational realisation

In this final chapter of the lecture we want to focus on how to implement variational regularisation methods numerically. We focus on variational regularisation methods as they appear to be the most general class of regularisation methods that we have studied throughout this lecture. However, we will only consider convex variational regularisation methods, as we want to guarantee that the iterative methods that we are going to consider converge to global minimisers.

4.1 A primal dual hybrid gradient method

Many Tikhonov-type regularisation problems of the form (3.26) can be formulated as a member of the following class of saddle-point problems:

$$\inf_u \sup_v F(u) - G(v) + \langle Du, v\rangle \,. \tag{4.1}$$

Here both F and G are proper, l.s.c. and convex functionals, and $D \in \mathcal{L}(\mathcal{U}, \mathcal{W})$. Before we continue to study (4.1) and potential numerical solutions, we want to give a quick example.

Example 4.1 (Total variation regularisation)**.** Note that we can reformulate (3.29) as

$$u_\alpha \in \operatorname*{arg\,min}_{u\in \mathrm{BV}(\Omega)} \left\{ \frac{1}{2}\|Ku - f\|_{\mathcal{V}}^2 + \sup_v \langle u, \alpha \mathrm{div} v\rangle - \chi_C(v) \right\}$$

with χ_C being a characterisitc function as defined in (3.31) and

$$C := \left\{ v \in C_0^\infty(\Omega; \mathbb{R}^n) \mid \| \|v\|_2 \|_{L^\infty(\Omega)} \leq 1 \right\}$$

In order to discuss how to solve (4.1) numerically, we want to consider the optimality conditions of the saddle-point problem first. For any subgradients $\hat{p} \in \partial F(\hat{u})$ and $\hat{q} \in \partial G(\hat{v})$ that satisfy

$$\begin{aligned} \hat{p} + D^*\hat{v} &= 0\,, \\ \hat{q} - D\hat{u} &= 0\,, \end{aligned} \tag{4.2}$$

the point $(\hat{u}, \hat{v})$ is a saddle-point of (4.1), due to the convexity of F and G. We therefore want to approach finding a saddle point of (4.1) via the iterative approach

$$\begin{pmatrix} p^{k+1} + D^*v^{k+1} \\ q^{k+1} - Du^{k+1} \end{pmatrix} + M \begin{pmatrix} u^{k+1} - u^k \\ v^{k+1} - v^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{4.3}$$

for $p^{k+1} \in \partial F(u^{k+1})$ and $q^{k+1} \in \partial G(v^{k+1})$, and a 2×2 operator-matrix M. We immediately observe that (4.3) should approach (4.2) in the limit (for properly chosen M). Hence, the important question is how to choose M such that (4.3) results in a convergent algorithm with relatively simple update steps for u^{k+1} and v^{k+1}? From our intuitive knowledge as numerical analysts we would guess that M should be positive definite and maybe even self-adjoint, to guarantee convergence of (4.3). But beyond that? A naïve choice for M could simply be the identity; however, in that case each update step is implicit in both u^{k+1} and v^{k+1}, which may not simplify solving our saddle-point problem at all. Alternatively, we propose to use

$$M := \begin{pmatrix} \frac{1}{\tau} I & -D^* \\ -D & \frac{1}{\sigma} I \end{pmatrix} \tag{4.4}$$

instead, with τ and σ being positive scalars. For this choice, (4.3) simplifies to the equations

$$u^{k+1} - (u^k - \tau D^* v^k) + \tau p^{k+1} = 0\,, \tag{4.5}$$

$$v^{k+1} - (v^k + \sigma D(2u^{k+1} - u^k)) + \sigma q^{k+1} = 0\,. \tag{4.6}$$

Due to $p^{k+1} \in \partial F(u^{k+1})$ and $q^{k+1} \in \partial G(u^{k+1})$, Equations (4.5) and (4.6) can be rewritten to

$$u^{k+1} = (I + \tau \partial F)^{-1}(u^k - \tau D^* v^k)\,, \tag{4.7}$$

$$v^{k+1} = (I + \sigma \partial G)^{-1}(v^k + \sigma D(2u^{k+1} - u^k))\,, \tag{4.8}$$

where $(I + \alpha \partial E)^{-1}$ is short for

$$(I + \alpha \partial E)^{-1}(f) := \arg\min_u \left\{ \frac{1}{2}\|u - f\|_2^2 + \alpha E(u) \right\}. \tag{4.9}$$

Operations of the form (4.9) are known as the *proximal* or *resolvent operator* with respect to the convex functional E.

The iterates (4.7) and (4.8) are known to be the so-called *primal-dual hybrid gradient method* (PDHGM). Before we prove actual convergence of those iterates to a saddle point of (4.1), we want to highlight what makes PDHGM so useful. First of all, the particular choice of M in (4.4) decouples u^{k+1} and v^{k+1} in the update for u^{k+1}, which makes updating u^{k+1} and v^{k+1} in an alternating possible. Secondly, PDHGM now only requires basic arithmetic operations, operator-adjoints and -multiplications, and the evaluation of the resolvent operations with respect to F and G. If these are simple, the overall PDHGM is simple. We want to take a look at the simple total variation problem to see that in that case the resolvent operations are particularly easy.

Example 4.2. Note that the resolvent operators in case of Example 4.1 are given as

$$\hat{u} = (I + \tau \partial F)^{-1}(w) = \arg\min_u \left\{ \frac{1}{2}\|u - w\|_2^2 + \frac{\tau}{2}\|Ku - f\|_2^2 \right\} \tag{4.10}$$

and

$$\hat{v} = (I + \sigma \partial G)^{-1}(z) = \arg\min_v \left\{ \frac{1}{2}\|v - z\|_F^2 + \sigma \chi_C(v) \right\}. \tag{4.11}$$

The resolvent operator (4.10) simply resembles the variational form of (shifted) Tikhonov regularisation. Hence, we can solve (4.10) via

$$\hat{u} = (I + \tau K^* K)^{-1}(w + \tau K^* f)\,,$$

according to Equation (3.11).

Further, we can figure out via case differentiation (left as an exercise) that the resolvent operator (4.11) can be computed via

$$\hat{v}_j = \frac{z_j}{\max(1, \|z\|_2)}\,,$$

where z_j denotes the j-th vectorial component of the vector field z, for $j \in \{1, \dots, n\}$. Hence, PDHGM reads as

$$u^{k+1} = (I + \tau K^* K)^{-1}(u^k + \tau(\mathrm{div} v^k + K^* f)) \tag{4.12}$$

$$v_j^{k+1} = \frac{(v^k + \sigma\nabla(2u^{k+1} - u^k))_j}{\max(1, \|v^k + \sigma\nabla(2u^{k+1} - u^k)\|_2)} \qquad \text{for all } j \in \{1, \dots, n\} \tag{4.13}$$

in case of total variation regularisation, due to $\nabla^* = -\mathrm{div}$. Note that in case of ROF-denoising as proposed in [8], Equation (4.12) simplifies to

$$u^{k+1} = \frac{u^k + \tau(\mathrm{div} v^k + f)}{1 + \tau}\,.$$

Note that for real world applications one obviously has to find appropriate discretisations of K and ∇.

4.1.1 Convergence of PDHGM

Before we prove convergence of the iterates under suitable conditions on τ and σ, we want to prove that M as defined in (4.4) is positive definite (under suitable conditions on τ and σ).

Lemma 4.1. *Consider the self-adjoint matrix M as defined in* (4.4)*. Then, M is positive definite if $0 < \tau\sigma\|D\|^2 < 1$.*

Proof. In order to prove positive definiteness of M we need to show that

$$\left\langle M \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\rangle > 0 \tag{4.14}$$

holds true for all $(u, v)^T \neq (0, 0)^T$. Evaluating the dual product in (4.14) yields the condition

$$\frac{1}{\tau}\|u\|^2 + \frac{1}{\sigma}\|v\|^2 - 2\langle Du, v\rangle > 0\,. \tag{4.15}$$

With the help of Young's inequality, we can estimate $-2\langle Du, v\rangle$ from below with

$$-2\langle Du, v\rangle \geq -\left(\tau\sigma\|D\|^2 \frac{c}{\tau}\|u\|^2 + \frac{1}{c\sigma}\|v\|^2\right)\,,$$

for some constant $c > 0$. Due to the condition $\tau\sigma\|D\|^2 < 1$ it is clear that we can find an $\varepsilon > 0$ with $(1 + \varepsilon)^2\tau\sigma\|D\|^2 = 1$. This further implies $1/((1 + \varepsilon)\tau\sigma\|D\|^2) = 1 + \varepsilon$ and hence, we can choose $c = 1 + \varepsilon = 1/((1 + \varepsilon)\tau\sigma\|D\|^2)$ to obtain

$$-2\langle Du, v\rangle \geq -\left(\frac{1}{1 + \varepsilon}\right)\left(\frac{1}{\tau}\|u\|^2 + \frac{1}{\sigma}\|v\|^2\right)\,.$$

Inserting this estimate into (4.15) therefore yields

$$\left(\frac{\varepsilon}{1+\varepsilon}\right)\left(\frac{1}{\tau}\|u\|^2+\frac{1}{\sigma}\|v\|^2\right)>0\,,$$

for $(u,v)^T \neq (0,0)^T$, since $\varepsilon > 0$. □

With the help of Lemma 4.1 we are able to prove convergence of the iterates (4.7) and (4.8) to a saddle point of (4.1).

Theorem 4.1. *Let $0 < \tau\sigma\|D\|^2 < 1$. Then the sequence (u^k, v^k) defined via (4.7) and (4.8) satisfies the following properties (for $\|u^0\| < \infty$ and $\|v^0\| < \infty$):*

- $\lim_{k\to\infty} D_F^{symm}(u^k, \hat{u}) = 0$
- $\lim_{k\to\infty} D_G^{symm}(v^k, \hat{v}) = 0$
- *All limit points of (u^k, v^k) are saddle points of (4.1).*

Here $(\hat{u}, \hat{v})$ denotes a saddle point satisfying (4.2).

Proof. Note that we can combine (4.2) and (4.3) with M as defined in (4.4) to

$$\begin{pmatrix} p^{k+1}-\hat{p}+D^*(v^{k+1}-\hat{v}) \\ q^{k+1}-\hat{q}-D(u^{k+1}-\hat{u}) \end{pmatrix}+\begin{pmatrix} \frac{1}{\tau}I & -D^* \\ -D & \frac{1}{\sigma}I \end{pmatrix}\begin{pmatrix} u^{k+1}-u^k \\ v^{k+1}-v^k \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{4.16}$$

Taking a dual product of

$$\begin{pmatrix} p^{k+1}-\hat{p}+D^*(v^{k+1}-\hat{v}) \\ q^{k+1}-\hat{q}-D(u^{k+1}-\hat{u}) \end{pmatrix}$$

with $(u^{k+1}-\hat{u}, v^{k+1}-\hat{v})^T$ yields

$$\left\langle \begin{pmatrix} p^{k+1}-\hat{p}+D^*(v^{k+1}-\hat{v}) \\ q^{k+1}-\hat{q}-D(u^{k+1}-\hat{u}) \end{pmatrix}, \begin{pmatrix} u^{k+1}-\hat{u} \\ v^{k+1}-\hat{v} \end{pmatrix} \right\rangle = D_F^{\mathrm{symm}}(u^{k+1},\hat{u})+D_G^{\mathrm{symm}}(v^{k+1},\hat{v}) \geq 0\,.$$

Hence, we can conclude

$$0 \geq \left\langle M\begin{pmatrix} u^{k+1}-u^k \\ v^{k+1}-v^k \end{pmatrix}, \begin{pmatrix} u^{k+1}-\hat{u} \\ v^{k+1}-\hat{v} \end{pmatrix}\right\rangle = \left\langle \begin{pmatrix} u^{k+1}-u^k \\ v^{k+1}-v^k \end{pmatrix}, M\begin{pmatrix} u^{k+1}-\hat{u} \\ v^{k+1}-\hat{v} \end{pmatrix}\right\rangle$$

due to M being self-adjoint. By using the notation $w^k := (u^k, v^k)^T$ and $\hat{w} := (\hat{u}, \hat{v})^T$, and $\|w\|_M := \sqrt{\langle M, w, w\rangle}$, we observe

$$\begin{aligned} 0 &\geq \|w^{k+1}-w^k\|_M^2 + \langle w^k-\hat{w}, M(w^{k+1}-w^k)\rangle \\ &= \|w^{k+1}-w^k\|_M^2 - \|w^k-\hat{w}\|_M^2 - \langle \hat{w}-w^k, M(w^{k+1}-\hat{w})\rangle \\ &= \|w^{k+1}-w^k\|_M^2 - \|w^k-\hat{w}\|_M^2 + \frac{1}{2}\left(\|w^{k+1}-\hat{w}\|_M^2+\|w^k-\hat{w}\|_M^2-\|w^{k+1}-w^k\|_M^2\right) \\ &= \frac{1}{2}\left(\|w^{k+1}-w^k\|_M^2 - \|w^k-\hat{w}\|_M^2+\|w^{k+1}-\hat{w}\|_M^2\right), \end{aligned}$$

and thus, we conclude

$$\|w^k-\hat{w}\|_M^2 \geq \|w^{k+1}-\hat{w}\|_M^2+\|w^{k+1}-w^k\|_M^2\,.$$

As $\|\cdot\|_M$ is a norm, due to M being self-adjoint and positive definite, this implies boundedness of the sequence w^k. Summing up the inner product of (4.16) with $w^{k+1} - \hat{w}$ over all k therefore yields

$$\begin{aligned}&\sum_{k=0}^{\infty} 2\left(D_F^{\text{symm}}(u^{k+1}, \hat{u}) + D_G^{\text{symm}}(v^{k+1}, \hat{v})\right) + \sum_{k=0}^{\infty} \|w^{k+1} - w^k\|_M^2 \\ = &\sum_{k=0}^{\infty} \left(\|w^k - \hat{w}\|_M^2 + \|w^{k+1} - \hat{w}\|_M^2\right) = \|w^0 - \hat{w}\|_M^2 < \infty\,.\end{aligned}$$

This implies $\|w^{k+1} - w^k\|_M \to 0$, $D_F^{\text{symm}}(u^{k+1}, \hat{u}) \to 0$ and $D_G^{\text{symm}}(v^{k+1}, \hat{v}) \to 0$ for $k \to +\infty$.

Due to the boundedness of w^k there exists a subsequence w^{k_l} that converges to a limit point $w^\infty := \lim_{l\to+\infty} w^{k_l}$. We have to verify that w^∞ is a saddle point, which for $s^\infty := \lim_{l\to+\infty} s^{k_l}$, $s^k := (p^k, q^k)^T$, follows immediately from (4.16). □

4.1.2 Deconvolution with total variation regularisation

An example for total variation regularisation of the inverse problem of image convolution is given as Exercise 2 on Exercisesheet 3.

Bibliography

[1] Heinz Werner Engl. *Integralgleichungen*. Springer-Verlag, 1997. 19

[2] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996. 16, 17, 19

[3] Charles W Groetsch. *Stable approximate evaluation of unbounded operators*. Springer, 2006. 23

[4] Jan Lellmann. Convex optimization with applications to image processing, 2013. Lecture notes. 36

[5] Terry J Morrison. *Functional analysis: An introduction to Banach space theory*, volume 43. John Wiley & Sons, 2011. 22

[6] Arch W Naylor and George R Sell. *Linear operator theory in engineering and science*. Springer Science & Business Media, 2000. 14

[7] R Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970. 36

[8] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 34, 41

[9] Terence Tao. *Epsilon of Room, One*, volume 1. American Mathematical Soc., 2010. 31

[10] Eberhard Zeidler. Applied functional analysis, applications to mathematical physics. *Applied Mathematical Sciences*, 108, 1995. 13, 14