Lappeenranta University of Technology
School of Engineering Science
Computational Engineering and Technical Physics
Technomathematics

Emelia Boye

REGULARIZATION OF INVERSE PROBLEMS BY THE LANDWEBER ITERATION

Master's Thesis

Examiners: Professor Tapio Helin
Professor Heikki Haario

Supervisor: Professor Tapio Helin

ABSTRACT

Lappeenranta University of Technology
School of Engineering Science
Computational Engineering and Technical Physics
Technomathematics

Emelia Boye

REGULARIZATION OF INVERSE PROBLEMS BY THE LANDWEBER ITERATION

Master's Thesis

2019

52 pages.

Examiners: Professor Tapio Helin
Professor Heikki Haario

Keywords: Ill-posed problems, regularization, inverse problems, Landweber iterative method, Discrepancy Principle, stopping rules

Landweber's method is a well-known iterative technique for regularizing linear and non-linear ill-posed equations. This thesis constructs in the Hilbert space setting a Landweber iteration to solve linear ill-posed inverse problems. Combined with an a-posteriori stopping principle known as the Discrepancy Principle, we show that the Landweber method is convergent. The fundamental principle of the criteria of Picard, the singular function analysis of Schmidt and, the concept of the generalized inverse of Moore-Penrose are illustrated.

PREFACE

Inverse problems arise in many cutting-edge systems where required data can only be retrieved from indirect measurements. Driven by domain needs, in the world of applied mathematics, the area of inverse problems has developed over the last two decades. This growth has been fostered by both advances in computation and theoretical breakthroughs. One challenge mathematicians have to address while dealing with inverse problems is the assumption that these problems are ill-posed in a mathematical sense, implying that quantities of study are not continuously dependent on the data and the cause of instability in their solutions under data perturbations. This problem requires the use of so-called regularization techniques and proven theories which has been developed. This thesis deals with the linear inverse problems described by the singular value expansion only.

More focus on the analysis of iterative regularization techniques has been investigated recently. Surprisingly, they turned out to be the best alternative approach to the famous Tikhonov regularization, especially for large scale inverse problems. This thesis is devoted to the convergence study of the Landweber iteration in the Hilbert setting derived during the last few years.

A number of excellent books on this subject have already appeared which the reader might refer to for aspects not fully treated here: [1–15].

My biggest appreciation is to my supervisor, Professor Tapio Helin for his immense unceasing contribution to the success of this thesis. His guidance helped me in all aspects of the research. I could not have asked for a better supervisor.

Words cannot express my gratitude to my family for their prayers and support. Finally, I appreciate all my fellow postgraduate students who helped in times of difficulties. I love you all.

Lappeenranta, November 15, 2019

Emelia Boye

CONTENTS

1 INTRODUCTION

1.1 Background

Making inferences with regards to physical parameters from measurements is a significant aspect of physical sciences. In particular, the physics laws provide the means to measure the data values given in a model. This is known as the *forward or direct problem*. *Direct* and *inverse problems* are opposites of each other. Generally, in inverse problems, one is given the effect and wishes to derive the cause. In contrast, direct problems seek "the effect given the cause". Let us consider a simple example that might serve as an illustration for the distinction of the two problems: Imagine a black box with an electrical circuit inside it and two knobs and a light bulb attached to the outside. By turning the knobs, we observe that the bulb produces different light intensities. Now, suppose the electrical circuit is known. Then with sufficient physics, we would be able to compute the light intensity that will result from a certain positioning of the two knobs. Doing this may be far from trivial. This is an example of a *direct problem* where the two knobs serve as the *cause* and the light intensity as the *effect*. A corresponding *inverse problem* is the following: Given the description of the circuit and observations of the light intensities, determine the corresponding positions of the two knobs.

According to Keller [16], two problems are called **inverses** of the other when the construction of one problem involves the solution of the other.

Typically, in inverse problems, one seeks to infer underlying parameters from observations and this is often an *ill-posed* problem. In the sense of Hadamard [17], a mathematical problem is said to be well-posed if it satisfies the following three properties or conditions:

1. **Existence:** There exists a solution to the problem.
2. **Uniqueness:** The solution is unique.
3. **Stability:** The solution depends continuously on the data.

A problem is considered ill-posed if any of the above three properties are violated and thus one has to use regularization methods as a remedy for the problem. Let us consider a simple linear inverse problem in Hilbert spaces to illustrate the fundamental concepts

of regularization. Our mathematical framework for the study of inverse problems is as follows:

Let H_1 and H_2 be two Hilbert spaces of finite or infinite dimensions and $A : H_1 \longrightarrow H_2$, a bounded linear operator on H_1, taking values in H_2. The problem is to find $x \in H_1$ which satisfies the equation

$$Ax = y \tag{1}$$

where $y \in H_2$ is given.

Let us consider equation (1) in the light of Hadamard criteria. We call y attainable or the existence of the solution is guaranteed if y belongs to the range $Ran(A)$ of A, where $Ran(A) := \{y \in H_2 \mid \text{ there exists } x \in H_1 : y = Ax\}$. The solution is unique if and only if $Ker(A) = \{0\}$, where $Ker(A) := \{x \in H_1 \mid Ax = 0\}$ denotes the kernel (null space) of A.

If property 1 and 2 hold, then $A^{-1} : Ran(A) \to H_1$ exists and property 3 is equivalent to the continuity (or boundedness) of A^{-1}. However, there is always a third obstacle for finding a useful solution since any practical observation is contaminated with noise. Instead of the exact equation (1), we observe

$$y^\delta = Ax + \delta,$$

where y^δ is the noisy data and δ is the noise. Also, if $Ker(A) \neq \{0\}$, which makes the solution of equation (1) not unique, one might be interested in a specific solution satisfying additional requirements.
Often in inverse problems, the operator A is compact. In such a case, if A^{-1} exists, it cannot be continuous unless the spaces $H_{j\in\{1,2\}}$ are finite dimensional. Thus, small noise in y can cause random size errors in x. The generalized notion of the solution will be provided via the concept of a *generalized inverse of A*; its continuity will then be relevant for property 3.

1.2 Inverse problems in general

Generally, inverse problems involve the task of "finding the cause" given the "*desired effect*". Concerning this aspect, one summarises terms into the following problems:

1. *Identification or reconstruction*, if the cause for an observed effect is sought for.

2. *Control or design,* if the cause of the desired effect is sought for.

Both of the above problems are related, but due to the different targets, there are several theoretical implications. For instance, in identifying a problem, the desired property is the uniqueness of the solution, since the observed effect is probably to have a specific cause one would like to receive. In a control problem, uniqueness is not important as non-uniqueness only means that different approaches can achieve the design goal.

Whereas the violation of Hadamard's property 1 is due to issues with the problem modelling, the violation of property 2 and 3 requires serious consideration. If a problem has more than one solution, one either has to decide which of these solutions is of interest (e.g the one with the smallest norm, which is appropriate for some, but not all, applications). One could compensate for this situation by feeding in additional information if possible. Violation of property 3, that is, the solution does not depend stably on the data, creates serious numerical problems: if one wants to approximate a problem whose solution does not depend continuously on the data by a traditional (standard) numerical method as one will use for a well-posed problem, then the numerical method becomes unstable. A partial remedy for this is the use of "*regularization methods*". This typically involves the inclusion of additional information in the data, such as smoothness of solution to obtain reasonable approximations for the ill-posed problem. A regularization method recovers partial information about the solution as stably as possible so as to find the right compromise between accuracy and stability. On the other hand, it is also good to have low computational expense of the algorithms, since in practice, the amount of information to be processed is voluminous.

There are numerous applications of inverse problems. A classical example is computerized tomography, where the forward map is modelled by a Radon transform operator. Other examples include image Deblurring and Denoising, Parameter Identification in dynamical systems, among others. Such examples are illustrated in [2, 18–20].

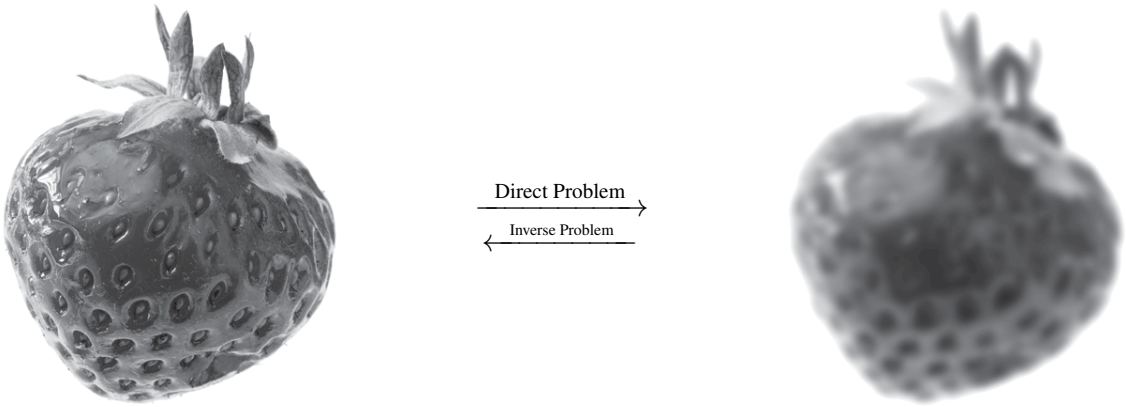
A graphical example of an inverse problem from image processing also known as *deblurring* is shown below. This type of inverse problem is to find the sharp photograph from a given blurry image.

In Figure 2, the cause is the sharp image (a) and the effect is the blurred image (b).

The main aim of this masters study is the regularization by an iterative method called the Landweber iteration of ill-posed inverse problems. Iterative techniques have been

(a) Computerized tomography (b) Image Deblurring (c) electrical impedance tomography

Figure 1. The graphical view of the some examples of inverse problem [21].



(a) Sharp image (b) Burred Image

Figure 2. The graphical view of deblurring effect in image processing [22].

successful in the sense that *regularized solutions* are achieved by stopping the processes at an early stage. On the other hand, the main challenge in the use of iterative methods is to choose the iteration's stopping index: stopping the iteration early produces an over-regularized solution, whilst a delay too produces a noisy solution.

1.3 Objectives and delimitations

1.3.1 Objectives

The main objective of this project is to produce a stable approximate solution of an ill-posed inverse problem by an iterative regularization theory which introduces prior knowl-

edge and make the approximation of this ill-posed inverse problem feasible, and to introduce some basic concepts of this theory by considering a very simple functional analytic framework known as compactness for operators acting on Hilbert spaces.

1.3.2 Delimitations of the Study

The main focus of thesis is the regularization theory in the Hilbert space setting for linear ill-posed problems. Also, one main iterative method namely the Landweber iteration is studied.

1.4 Structure of the thesis

This thesis is organised into four chapters. Chapter 1 provides an introduction to the thesis which includes some background study of inverse problems, objectives and delimitations of the study. Chapter 2 provides the preliminaries to the thesis. Existing theories that are relevant to the study are reviewed and formulas derived. We recall the fundamental notions of regularization in Hilbert space setting. We will follow the famous book of Engl, Hanke and Neubauer [2], the lecture notes of Burger [19] and the research work of Tomba [20]. Chapter 3 describes in details regularization theory in linear inverse problems and the regularization methods. Here, the various types of stopping rules are analysed. In Chapter 4, the main results of the thesis known as the Landweber iteration is presented and the an a-posteriori parameter choice rule known as the Discrepancy Principle of Morozov is also treated here.

2 PRELIMINARIES

There were some illustrations of ill-posed problems in the previous section. From Hadamard's definition of ill-posedness, it is obvious to adopt an extensive mathematical concepts to such problems which are based on a different concept of solution to the equation (1) to achieve the uniqueness and existence properties. The linear problems in the Hilbert space framework employs the Moore-Penrose Generalized Inverses.
Let us firstly give some known definitions and notations. Our main references for this section are [1, 2, 4, 19, 20, 22–35].

2.1 Fundamental Tools In the Hilbert Space Setting

Let A be a bounded linear operator between two Hilbert spaces H_1 and H_2 with the scalar products and their induced norms in H_1 and H_2 denoted $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. For $\bar{x} \in H_1$ and $\delta > 0$, we write

$$\mathcal{B}_\delta(\bar{x}) := \{x \in H_1 \mid \ \|x - \bar{x}\| < \delta\} \tag{2}$$

for the open ball centered in $\bar{x}$ with radius δ and we write $\overline{\mathcal{B}_\delta(\bar{x})}$ for its closure with respect to the topology of H_1.

Let $\mathcal{D}(A)$ denote the domain of A. Recall that the null space $Ker(A) \subset H_1$ is closed. The following definitions are fundamental.

Definition 2.1 (Orthogonal Space)**.** *Let $\mathcal{M} \subseteq H_1$. The orthogonal space of $\mathcal{M}$ is the closed subspace $\mathcal{M}^\perp$ defined by:*

$$\mathcal{M}^\perp := \{x \in H_1 \mid \langle x, z \rangle = 0, \ \ \textit{for all} \ \ z \in \mathcal{M}\}. \tag{3}$$

If $\mathcal{M}$ is also closed, then H_1 is the direct sum of $\mathcal{M}$ and $\mathcal{M}^\perp$ and we write $H_1 = \mathcal{M} \oplus \mathcal{M}^\perp$.

Definition 2.2 (Adjoint operator)**.** *The bounded operator $A^* : H_2 \to H_1$, defined as*

$$\langle A^* y, x \rangle = \langle y, Ax \rangle, \qquad \textit{for all} \ \ x \in H_1 \ \ \textit{and} \ \ y \in H_2, \tag{4}$$

is called the adjoint operator of A. If $A : H_1 \to H_1$ and $A = A^*$, then A is called self-adjoint.

Definition 2.3 (Closed linear operator)**.** *Let H_1 and H_2 be two Hilbert spaces such that $A : \mathcal{D}(A) \longrightarrow H_2$ a linear operator with domain $\mathcal{D}(A) \subset H_1$. Then A is called a closed linear operator if its graph*

$$gr(A) = \{(x, y) \mid x \in \mathcal{D}(A),\ \ y = Ax\}$$

is closed in the space $H_1 \times H_2$, where the two algebraic operations of the inner product in $H_1 \times H_2$ are defined as usual and the norm on $H_1 \times H_2$ is defined by

$$\|\langle x, y\rangle\| = (\|x\|^2 + \|y\|^2)^{\frac{1}{2}}. \tag{5}$$

Definition 2.4 (Orthogonal Projector)**.** *Let $\mathcal{W} \subset H_1$ be a closed subspace. For any $x \in H_1$, there exists a unique element $w \in \mathcal{W}$, called the projection of x onto $\mathcal{W}$, that minimizes the distance $\|x - w\|$, simultaneously for all $w \in \mathcal{W}$. The map $P_{\mathcal{W}} : H_1 \to H_1$, that associates to an element $x \in H_1$ its projection onto $\mathcal{W}$, is called the orthogonal projector onto $\mathcal{W}$ such that $x - w \in \mathcal{W}^{\perp}$. This is the unique linear and bounded self-adjoint operator that maps H_1 onto $\mathcal{W}$ such that $P_{\mathcal{W}} = P_{\mathcal{W}}^2$ and $\|P_{\mathcal{W}}\| = 1$.*

Definition 2.5. *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in H_1, $x \in H_1$. The sequence x_n is said to weakly converge to x if for all $z \in H_1$, $\langle x_n, z\rangle$ converges to $\langle x, z\rangle$. We denote weak convergence by*

$$x_n \rightharpoonup x \quad \textit{as} \ \ n \to \infty.$$

2.2 The Moore-Penrose Generalized Inverse

Consider equation (1) in Euclidean spaces: Computing the generalised inverse when A is a matrix of full rank is relatively easy. Also, if $Ran(A)$ is not the full image space, then the right hand side y of equation (1) becomes complicated to solve. In such a case, it is important to find x such that Ax has the minimal distance to x. If on the other hand $Ker(A) \neq \{0\}$, then equation (1) does not have a unique solution and one might be interested in choosing the specific solution with minimal norm among the multiple solutions.

In the Hilbert space real-valued setting, we consider following definition:

Definition 2.6. *Let $A : H_1 \to H_2$ be bounded linear operator. An element $x \in H_1$ is called*

(i) least squares solution of equation (1) *if*

$$\|Ax - y\| = \inf\{\|Az - y\| \ \ \mid z \in H_1\}. \tag{6}$$

(ii) best-approximate solution or minimum norm solution of equation (1) *if x is a least square solution of equation* (1) *and*

$$\|x\| = \inf\{\|z\| \quad | \ z \ \textit{is least-squares solution of equation } (1)\}. \tag{7}$$

It is easily seen that if the least squares solution exists, then the minimum norm solution is unique because it is the minimizer of quadratic error functions and thus the best-approximate can be defined as the least-squares solution of minimal norm. The notion of a best-approximate solution is related to the Moore-Penrose generalized inverse of A.

Given bounded linear operators A only, that is, $A \in \mathcal{L}(H_1, H_2)$ where $\tilde{A} : Ker(A)^{\perp} \to Ran(A)$ is its restriction, then *the Moore-Penrose generalised inverse* of A, denoted $A^{\dagger}$, is the unique linear extension of $\tilde{A}^{-1}$ to the domain of $A^{\dagger}$ denoted as:

$$\mathcal{D}(A^{\dagger}) := Ran(A) \oplus Ran(A)^{\perp} \tag{8}$$

such that

$$Ker(A^{\dagger}) = Ran(A)^{\perp}. \tag{9}$$

Thus due to the restriction to $Ker(A)^{\perp}$, $\tilde{A}$ is injective (one-to one) and surjective (onto) due to the restriction to $Ran(A)$. Hence, $\tilde{A}$ is bijective, and, $\tilde{A}^{-1}$ exists. For any $y \in \mathcal{D}(A^{\dagger})$, there is unique $y_1 \in Ran(A)$ and $y_2 \in Ran(A)^{\perp}$ with $y = y_1 + y_2$. Since $Ker(\tilde{A}) = \{0\}$ and $Ran(\tilde{A}) = Ran(A)$, the operator $A^{\dagger}$ is well-defined and from (9) and the linearity of $A^{\dagger}$, we have that

$$A^{\dagger}y = A^{\dagger}y_1 + A^{\dagger}y_2 = A^{\dagger}y_1 = \tilde{A}^{-1}y_1. \tag{10}$$

Proposition 2.7. *Let $P = AA^{\dagger}$ and $Q = A^{\dagger}A$ be the orthogonal projection operators onto $Ker(A)$ and $\overline{Ran(A)}$, respectively. Then $A^{\dagger}$ is uniquely characterized by the four Moore-Penrose equations:*

$$\begin{aligned} AA^{\dagger}A &= A && (11)\\ A^{\dagger}AA^{\dagger} &= A^{\dagger} && (12)\\ A^{\dagger}A &= I - P && (13)\\ AA^{\dagger} &= Q\,|_{D(A^{\dagger})}, && (14) \end{aligned}$$

where I is the identity operator.

Proof. For each $y \in \mathcal{D}(A^{\dagger})$ and by the definition of the Moore-Penrose inverse $A^{\dagger}$, we

have

$$A^\dagger y = \tilde{A}^{-1} Q y = A^\dagger Q y \tag{15}$$

so that $A^\dagger y \in Ran(\tilde{A}^{-1}) = Ker(A)^\perp$. For each $x \in Ker(A)^\perp$, it follows that

$$A^\dagger A x = \tilde{A}^{-1} \tilde{A} x = x.$$

The above assertion proves that $Ran(A^\dagger) = Ker(A)^\perp$. Now equation (15) implies that

$$A A^\dagger y = A A^\dagger Q y = A \tilde{A}^{-1} Q y = \tilde{A} \tilde{A}^{-1} Q y = Q y,$$

since $\tilde{A}^{-1} Q y \in Ker(A)^\perp$ and hence equation (14) holds. By the definition of $A^\dagger$, it holds for each $x \in H_1$ that

$$A^\dagger A x = \tilde{A}^{-1} A \left(P x + (I - P) x \right) = \tilde{A}^{-1} \underbrace{A P x}_{=0} + \tilde{A}^{-1} \tilde{A} (I - P) x = (I - P) x. \tag{16}$$

Inserting equation (14) into equation (15) yields

$$A^\dagger y = A^\dagger Q y = A^\dagger A A^\dagger y$$

for all $y \in D(A^\dagger)$. equation (16) implies equation (13) and equation (13) also implies that

$$A A^\dagger A = A(I - P) = A - AP = A.$$

Hence equations (11), (15) and (14) imply equation (12). □

Any operator satisfying equation (13) or equation (14) is referred to as an *inner inverse* or *outer inverse* of A, respectively.
The following theorem provides a connection between the least-squares solutions and how they are computed via the Moore-Penrose inverse.

Theorem 2.8. *For all $y \in \mathcal{D}(A^\dagger)$, the equation* (1) *has a unique best-approximate solution given by*

$$x^\dagger := A^\dagger y.$$

The set of all least-squares solution is $A^\dagger y + Ker(A)$.

Proof. For a fixed $y \in \mathcal{D}(A^\dagger)$, let us construct a set

$$S = \{ z \in H_1 \mid A z = Q y \}.$$

Since $y \in \mathcal{D}(A^\dagger) = Ran(A) \oplus Ran(A)^\perp$, it follows that $Qy \in Ran(A)$ and therefore $S \neq \emptyset$. Because Q is an orthogonal projector we have for all $z \in S$ and for all $x \in H_1$:

$$\|Az - y\| = \|Qy - y\| \leq \|Ax - y\|.$$

So, all elements in S are least-squares solutions of $Ax = y$. Conversely, let z be a least-squares solutions of $Ax = y$. Then

$$\|Qy - y\| \leq \|Az - y\| = \inf\{\|u - y\| \mid u \in Ran(A)\} = \|Qy - y\|.$$

Thus, Az is the closest element to y in $Ran(A)$, that is, $Az = Qy$ and

$$S = \{x \in H_1 \mid x \text{ is least-squares solution of } Ax = y\}.$$

Now, let $\bar{z}$ be the element of minimal norm in $S = A^{-1}(\{Qy\})$. Since then $S = \bar{z} + Ker(A)$, it suffices to show that

$$\bar{z} = A^\dagger y.$$

As an element of minimal norm in $S = \bar{z} + Ker(A)$, $\bar{z}$ is orthogonal to $Ker(A)$, that is $\bar{z} \in Ker(A)^\perp$. This implies that

$$\bar{z} = (I - P)\bar{z} = A^\dagger A\bar{z} = A^\dagger Qy = A^\dagger AA^\dagger y = A^\dagger y, \tag{17}$$

that is, $\bar{z} = A^\dagger y$. □

In linear algebra, it is a well-known fact that the least-squares solutions can be characterized by the normal equations and this leads us to the next theorem to verify if the assertion is true in the continuous case.

Theorem 2.9. *Let A^* denote the adjoint operator of A. For given $y \in \mathcal{D}(A^\dagger)$, $x \in H_1$ is a least-squares solution of equation* (1) *if and only if x satisfies the normal equations*

$$A^*Ax = A^*y. \tag{18}$$

Proof. An element $x \in H_1$ is least-squares solution of equation (1) if and only if Ax coincides with the projection of y onto $Ran(A)$. This is equivalent to $Ax - y \in Ran(A)^\perp = Ker(A^*)$. Thus, we conclude that $A^*(Ax - y) = 0$ which is equivalent to equation (18). Furthermore, a least-squares solution has minimal norm if and only if $x \in Ker(A)^\perp$. □

A direct consequence from Theorem 2.9 is that $A^\dagger y$ is a solution of equation (18) with

minimal norm. Consequently, we have

$$A^{\dagger}y = (A^{*}A)^{\dagger}A^{*}y, \tag{19}$$

and this means that in order to approximate $A^{\dagger}y$, we may be required to compute an approximation via equation (18) instead.

To analyse the domain of the generalised inverse, one could show from the Moore-Penrose inverse as defined earlier in equation (8) that $\mathcal{D}(A^{\dagger})$ is the natural domain of the definition for $A^{\dagger}$. This is in the sense that if $y \notin \mathcal{D}(A^{\dagger})$, then there is no existence of least-squares solution of equation (1). Thus contrary to the finite-dimensional case, the concept of minimal-norm solution as introduced does not always give a solution to a problem although the concept imposes uniqueness. As we know A is a bounded linear operator and thus the orthogonal complements always admits a closure. We can then conclude from equation (8) that

$$\overline{\mathcal{D}(A^{\dagger})} := \overline{Ran(A)} \oplus Ran(A)^{\perp} = \overline{Ran(A)} \oplus Ker(A^{*})^{\perp} = H_2.$$

The domain $\mathcal{D}(A)$ is dense in H_1 whereas $\mathcal{D}(A^{\dagger})$ is dense in H_2. Thus, it follows that $\mathcal{D}(A^{\dagger}) = H_2$ if $Ran(A)$ is closed and vice versa. $\mathcal{D}(A^{\dagger}) = H_2$ implies that $Ran(A)$ is also closed. Furthermore, for $y \in Ran(A)^{\perp} = Ker(A^{\dagger})$, the best-approximate solution is $x^{\dagger} = 0$. It is therefore important to check when y also satisfies $y \in Ran(A)$, for any given $y \in \overline{Ran(A)}$. In such a case, $A^{\dagger}$ has to be continuous. However, if $y \in \overline{Ran(A)} \setminus Ran(A)$ exists, then it is just enough to prove that $A^{\dagger}$ is discontinuous. This leads us to the following theorems and the introduction of compact operators which discuss the discontinuities of the Moore-Penrose inverses.
The following theorem is a result which implies that the $Ran(A)$ of continuous operator between two Hilbert spaces is closed.

Theorem 2.10 (Bounded inverse theorem)**.** *Let H_1 and H_2 be Hilbert spaces. If $A \in \mathcal{L}(H_1, H_2)$ is bijective (one-to-one and onto mapping), then the inverse map $A^{-1} \in \mathcal{L}(H_1, H_2)$.*

Proof. The proof of Theorem 2.10 can be found in [28, Theorem 8.72]. □

The Closed Graph Theorem provides conditions for which a closed linear operator as defined in Definition 2.3 is bounded.

Theorem 2.11 (Closed Graph Theorem)**.** *Let H_1 and H_2 be Hilbert spaces and let A :*

$\mathcal{D}(A) \to H_2$ be a linear operator from H_1 to H_2. Then if $A : \mathcal{D}(A) \to H_2$ is a closed linear operator and its domain $\mathcal{D}(A)$ is closed in H_1, then the operator A is bounded.

Proof. Assume that $H_1 \times H_2$ is complete. Assume also that $gr(A)$ is a closed subspace in $H_1 \times H_2$ and $\mathcal{D}(A)$ is a closed subspace in H_1, thus $gr(A)$ and $\mathcal{D}(A)$ are complete. We now define the projection mapping

$$P : gr(A) \longrightarrow \mathcal{D}(A)$$

by

$$P(x, Ax) := x.$$

The mapping P is linear and bounded since

$$\|P(x, Ax)\| = \|x\| \leq \|x\| + \|Ax\| = \|(x, Ax)\|$$

for all $x \in \mathcal{D}(A)$. In fact, its inverse

$$P^{-1} : \mathcal{D}(A) \to gr(A)$$

is defined by

$$P^{-1}x := (x, Ax).$$

for all $x \in H_1$. Since $gr(A)$ and $\mathcal{D}(A)$ are complete, by the bounded inverse theorem 2.10, the projection P^{-1} is bounded and there is a constant b such that

$$\|(x, Ax)\| = \|P^{-1}x\| \leq b\|x\|$$

for all $x \in \mathcal{D}(A)$. But this implies that A is bounded since

$$\|Ax\| \leq \|Ax\| + \|x\| = \|(x, Ax)\| \leq b\|x\|$$

for all $x \in \mathcal{D}(A)$. □

Closed Graph Theorem is applied to Moore-Penrose generalized inverse and gives the following result.

Theorem 2.12. *Let $A \in \mathcal{L}(H_1, H_2)$. Then $A^\dagger \in \mathcal{L}(\mathcal{D}(A^\dagger), H_1)$ if and only if $Ran(A)$ is closed.*

Proof. Before proving the result, we will derive the following identity:

$$\{(y_1, \tilde{A}^{-1}y_1) \mid y_1 \in Ran(A)\} = \{(Ax, x) \mid x \in H_1\} \cap (H_2 \times Ker(A)^{\perp}). \tag{20}$$

Let $y_1 \in Ran(A)$, $x := \tilde{A}^{-1}y_1$; by the definition of $\tilde{A}$, $x \in Ker(A)^{\perp}$, and due to equation (14), we have

$$Ax = AA^{\dagger}y_1 = y_1.$$

Hence, it follows that

$$(y_1, \tilde{A}^{-1}) = (Ax, x) \in H_2 \times Ker(A)^{\perp}.$$

If $x \in Ker(A)^{\perp}$ and $y_1 := Ax$ (hence, $y_1 \in Ran(A)$), then $\tilde{A}^{-1}y_1 = A^{\dagger}Ax = x$, so that

$$(y_1, \tilde{A}^{-1}y_1) = (Ax, x).$$

Thus, equation (20) holds.
By the definition of $A^{\dagger}$, we have for the graph of $A^{\dagger}$:

$$\begin{aligned} gr(A^{\dagger}) &= \{(y, A^{\dagger}y) \mid y \in \mathcal{D}(A^{\dagger})\} \\ &= \{(y_1 + y_2, \tilde{A}^{-1}y_1) \mid y_1 \in Ran(A), y_2 \in Ran(A)^{\perp}\} \\ &= \{(y_1, \tilde{A}^{-1}y_1) \mid y_1 \in Ran(A)\} + (Ran(A)^{\perp} \times \{0\}), \end{aligned}$$

which implies with equation (20) that

$$gr(A^{\dagger}) = [\{(Ax, x) \mid x \in H_1\} \cap (H_2 \times Ker(A)^{\perp})] + [Ran(A)^{\perp} \times \{0\}]. \tag{21}$$

The spaces on the right-hand side of equation (21) are closed and orthogonal to each other in $H_2 \times H_1$, so that also their sum $gr(A^{\dagger})$ is also closed.
To prove the second part of the proposition, we assume now that $Ran(A)$ is closed, so that $\mathcal{D}(A^{\dagger}) = H_2$. Because of the Closed Graph Theorem 2.11, $A^{\dagger}$ is bounded. Conversely, let $A^{\dagger}$ be bounded, then $A^{\dagger}$ has a unique continuous extension $\overline{A^{\dagger}}$ to H_2. From equation (14) and the continuity of A we conclude that

$$A\overline{A^{\dagger}} = Q.$$

Hence for $y \in \overline{Ran(A)}$, we have $y = Qy = A\overline{A^{\dagger}}y \in Ran(A)$. Consequently, we obtain

$$\overline{Ran(A)} \subseteq Ran(A)$$

and therefore $Ran(A)$ is closed. □

2.3 Compact Operators

Let us introduce the concept of compactness for linear operators.

Definition 2.13 (Compact Linear Operator)**.** *Let H_1 and H_2 be Hilbert spaces. Let $A \in \mathcal{L}(H_1, H_2)$. Then the operator $A : H_1 \to H_2$ is said to be compact or a completely continuous linear operator if for every bounded subset of M of H_1, $M \subset H_1$, the image $A(M) \subset H_2$ is relatively compact, that is, has compact closure, $\overline{A(M)}$. Alternatively, A is said to be compact if the image of every bounded sequence $(x_n)_{n=1}^{\infty}$ in H_1, the sequence $(Ax_n)_{n=1}^{\infty}$ in H_2 contains a converging subsequence.*

Later, we study inverse problems that involve compact operators. As we will notice, the compactness of the forward operator is the source of ill-posedness in equation (1). This is demonstrated by the following theorem.

Theorem 2.14. *Let $A : H_1 \to H_2$ be the space of compact linear operators between infinite dimensional Hilbert spaces H_1 and H_2, such that the dimension of $Ran(A)$ is infinite. Then the problem in equation* (1) *is ill-posed, that is $A^\dagger$, the Moore-Penrose inverse is discontinuous.*

Proof. The space H_1 and $Ran(A)$ are infinite dimensional which also implies that $Ker(A)^\perp$ is infinite dimensional. Note however that the dimension of $Ran(A)$ is always less or equal to the dimension of $Ker(A)^\perp$. Therefore, one can find a sequence $\{x_n\}_{n=1}^{\infty} \subset Ker(A)^\perp$ with $\|x_n\| = 1$ such that

$$\langle x_n, x_k \rangle = 0 \text{ for } n \neq k.$$

Since A is a compact operator, the sequence $\{y_n := Ax_n\} \subset Ran(A)$ is compact. Hence, one can find k, ℓ for each $\delta > 0$ such that $\|Ax_k - Ax_\ell\| = \|y_k - y_\ell\| < \delta$. However,

$$\|A^\dagger y_k - A^\dagger y_\ell\|^2 = \|x_k - x_\ell\|^2 = \|x_k\|^2 + \|x_\ell\|^2 - 2\langle x_k, x_\ell \rangle = 2$$

Hence, $A^\dagger$ is unbounded and thus, discontinuous. □

Definition 2.15. *If $A \in \mathcal{L}(H_1, H_2)$ and $Ran(A)$ is finite-dimensional, then the operator A has **finite rank**.*

2.4 Singular Value Expansion

The common way to prove the compactness of an operator is to approximate the operator by other operators, for example, by finite rank operators which are compact. Using such an approximate scheme, the following results are usually employed.

Theorem 2.16. *Let $A : H_1 \to H_2$ be a linear operator and $A_n \in \mathcal{L}(H_1, H_2)$ a sequence of compact operators. If $A_n \to A$ in the operator norm, then A is a compact operator.*

Proof. Using the diagonalization sequence argument, we show that for a given bounded sequence $\{x_n\}_{n=1}^{\infty} \subset H_1$, the image $\{Ax_n\}_{n=1}^{\infty}$ has a convergent (hence Cauchy) subsequence. Since A_1 is a compact operator, we know that the sequence A_1x_n has a convergent subsequence $\{A_1x_{1,n}\}_{n=1}^{\infty}$. Similarly, since the subsequence $\{x_{1,n}\}_{n=1}^{\infty}$ is bounded and A_2 is compact, we can repeat the argument with A_2 to produce a subsequence $\{x_{2,n}\}_{n=1}^{\infty}$ of $\{x_{1,n}\}_{n=1}^{\infty}$ and we see that $\{A_2x_{1,n}\}_{n=1}^{\infty}$ has a convergent subsequence $\{A_2x_{2,n}\}_{n=1}^{\infty}$. We now repeat the argument, taking further subsequences of subsequences so that $\{x_{k,n}\}_{n=1}^{\infty}$ is a subsequence of $\{x_{j,n}\}_{n=1}^{\infty}$ if $j < k$ and so that $\{A_kx_{k,n}\}_{n=1}^{\infty}$ converges. We already know that $\{A_kx_{k,n}\}_{n=1}^{\infty}$ is convergent and hence it is Cauchy.
We continue in the same way and now consider the diagonal sequence $\{x_{n,n}\}_{n=1}^{\infty}$. We define a sequence $y_n := x_{n,n}$. Note that $\{y_n\}_{n=1}^{\infty}$ is a subsequence of the original sequence $\{x_n\}_{n=1}^{\infty}$ such that for every fixed positive integer m, the sequence A_my_n is convergent and hence Cauchy. The sequence $\{x_n\}_{n=1}^{\infty}$ is bounded, say by $\|x_n\| \leq c$, for all n. Hence $\|y_n\| \leq c$, for all n .
We claim that $\{Ay_n\}_{n=1}^{\infty}$ is Cauchy sequence in H_2. Let $\epsilon > 0$ be given. Since $A_n \to A$, there exists $p \in \mathbb{N}$ such that

$$\|A_p - A\| < \frac{\epsilon}{3c}. \tag{22}$$

Also, since $\{A_py_n\}_{n=1}^{\infty}$ is Cauchy. This is true since $\{y_n\}_{n=p}^{\infty}$ is a subsequence of $\{x_{p,m}\}_{m=1}^{\infty}$. Thus, there is an $N > 0$ such that

$$\|A_py_j - A_py_k\| < \frac{\epsilon}{3} \tag{23}$$

for all $j, k > N$. Therefore, for $j, k > N$, we have

$$\begin{aligned}
\|Ay_j - Ay_k\| &\leq \|Ay_j - A_py_j + A_py_j - A_py_k + A_py_k - Ay_k\| \\
&\leq \|Ay_j - A_py_j\| + \|A_py_j - A_py_k\| + \|A_py_k - Ay_k\| \\
&< \|A - A_p\|\|y_j\| + \frac{\epsilon}{3} + \|A - A_p\|\|y_k\| \\
&< \frac{\epsilon}{3c}c + \frac{\epsilon}{3} + \frac{\epsilon}{3c}c = \epsilon.
\end{aligned}$$

This proves that Ay_n is Cauchy and converges since H_2 is complete by definition. We have thus produced, for an arbitrary bounded sequence $(x_n) \subset H_1$, a convergent subsequence of its image under the operator A. Therefore, A is compact. □

Example 2.17. *Let $\Omega \subseteq \mathbb{R}^n$ be a bounded, open set and $(s,t) \in \Omega \times \Omega$. A kernel k is said to be weakly singular if and only if k is smooth such that $s \neq t$ and there exist constants $N > 0$ and $\nu < n$ for every $s \neq t \in \Omega$ such that*

$$|k(s,t)| \leq N|s-t|^{-\nu}.$$

Let

$$(Ax)(s) = \int_{\Omega} k(s,t)x(t)dt \tag{24}$$

where $x(t)$ is unknown and we wish to find with $k \in L^2(\Omega \times \Omega)$ (i.e., $k : \Omega \times \Omega \to \mathbb{R}$ is Hilbert-Schmidt kernel) or weakly singular. Then $A \in \mathcal{L}(L^2(\Omega), L^2(\Omega))$ is compact.

An analysis of the solutions to equation (1) can be based on the singular value expansion (SVE) of the kernel k. The SVE was developed by Schmidt [36].

If A is a continuous linear operator from Hilbert spaces H_1 to H_2, then the adjoint of A is denoted A^* such that $A^* : H_2 \to H_1$ is the continuous linear operator defined by

$$\langle Ax, y\rangle = \langle x, A^*y\rangle$$

for all $x \in H_1$ and $y \in H_2$. As a consequence of using the definitions of $Ran(A)$ and $Ker(A)$, we have the following result.

Theorem 2.18. *If $A : H_1 \to H_2$ is a continuous linear operator, then*

$$Ran(A)^{\perp} = Ker(A^*) \text{ and } Ker(A)^{\perp} = \overline{Ran(A^*)}.$$

Proof. The proof of Theorem 2.18 is available in [4]. □

The spectrum of a linear operator $A : H \to H$ denoted $\sigma(A)$ is the set of complex numbers defined by

$$\sigma(A) = \{\lambda \in \mathbb{C} : A - \lambda I \text{ does not have an inverse that is bounded}\},$$

where I is the identity operator on H. The spectral radius of A is the real number $|\sigma(A)|$

defined by

$$|\sigma(A)| = \sup\{|\lambda| : \lambda \in \sigma(A)\}.$$

As A is assumed to be bounded, it means $\sigma(A)$ is closed and $|\sigma(A)| \leq \|A\|$, and therefore $\sigma(A)$ is compact. The spectrum $\sigma(A)$ is a non-empty set of real numbers if A is a bounded self adjoint linear operator.

Compact linear self-adjoint operators have a simple spectrum with a notion of an eigensystem which plays an important role. Every non-zero member of the spectrum is an isolated point which is an eigenvalue of these operators. For every non-zero eigenvalue λ of a compact linear self-adjoint operator A, the eigenspace associated with the eigenvalue λ, that is, the set $Ker(A - \lambda I)$, is finite-dimensional and the eigenvalues form a sequence $\lambda_1, \lambda_2, \cdots$, which (if is infinite) approaches zero. The dimension of $Ker(A-\lambda I)$ is called its *multiplicity*. Every eigenvalue according to the dimension of its associated eigenspace when repeated can form a sequence $v_1, v_2, \cdots$, of associated orthonormal eigenvectors. This leads us to the spectral representation theorem for a compact linear self-adjoint operator.

Theorem 2.19. *Suppose $A : H_1 \to H_2$ is a compact linear self-adjoint operator, then the spectrum of A is given by $\{0\} \cup \{\lambda_n\}_{n=1}^{\infty}$ and the eigensystem $\{\lambda_n, v_n\}_{n=1}^{\infty}$ of A consists of non-zero real eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ (at most countably many which are repeated according to the dimension of the associated eigenspace) and a corresponding orthonormal eigenvectors $\{v_n\}_{n=1}^{\infty}$ (a complete orthogonal and normalized eigenvectors). The sequence $\{\lambda_n\}_{n=1}^{\infty}$ has non-zero accumulation point. For all $x \in H_1$, A can be diagonalized as*

$$Ax = \sum_{n=1}^{\infty} \lambda_n \langle x, v_n \rangle v_n \tag{25}$$

for all $x \in H_1$. Moreover, $Av_n = \lambda_n v_n$.

Proof. The first part of the proof is to show that the sequence of eigenvalues has no non-zero accumulation point. The proof is trivial if $A = 0$. In the non-trivial sense, we assume $\|A\| \neq 0$. Since A is self-adjoint, it follows that there exists

$$\alpha(A) = \sup_{\|x\|_{H_1}=1} |\langle x, Ax \rangle| = \|A\|. \tag{26}$$

Let $\{x_k\}_{k=1}^{\infty} \subset H_1$ with $\|x_k\| = 1$ for all $k \in \mathbb{N}$ be defined by

$$|\langle x_k, Ax_k \rangle| \to \alpha(A) = \|A\|, \quad k \to \infty.$$

The scalar products are all real with accumulation points $\pm\|A\|$. By considering λ_1 as a

subsequence of λ_n, we can assume without loss of generality that

$$\langle x_k, Ax_k \rangle \to \lambda_1, \quad k \to \infty, \tag{27}$$

where $\lambda_1 \in \{\pm\|A\|\}$. The sequence $(x_k)_k$ is bounded due to the compactness of A and hence there is a subsequence such that $v_1 \in H_1$ with

$$x_k \rightharpoonup v_1 \text{ and } Ax_k \to Av_1, \quad k \to \infty.$$

From equation (27), we have

$$\begin{aligned} \|Ax_k - \lambda_1 x_k\|^2 &= \|Ax_k\|^2 - 2\lambda_1 \langle x_k, Ax_k \rangle + \lambda_1^2 \|x_k\|^2 \\ &\leq \|A\|^2 - 2\lambda_1 \langle x_k, Ax_k \rangle + \lambda_1^2 \\ &= 2\lambda_1^2 (\lambda_1 - \langle x_k, Ax_k \rangle) \end{aligned}$$

converges to zero as $k \to \infty$. It follows that

$$\lambda_1 x_k = (\lambda_1 x_k - Ax_k) + (Ax_k - Av_1) + Av_1 \to Av_1$$

as $k \to \infty$. Since $x_k \rightharpoonup v_1$, we have that v_1 is a unit norm eigenfunction of A for λ_1. Now, we let $V_1 = span\{v_1\}^\perp \subset H_1$. We observe that

$$\langle Ax, v_1 \rangle = \langle x, Av_1 \rangle = \lambda_1 \langle x, v_1 \rangle = 0$$

for all $x \in V_1$. The subspace V_1 is an invariant for A. We can thus define

$$A_2 = A|_{V_1}$$

where A_2 belongs to the compact self-adjoint operators in V_1, and since $V_1 \subset H_1$, the corresponding norm of A_2 is also bounded by $|\lambda_1| = \|A\|$. In the case of $A_2 = 0$, the proof is trivial, otherwise, we can proceed as before and find $\lambda_2 \in \mathbb{R}$ and a corresponding $v_2 \in V_1 \subset H_1$, such that

$$Av_2 = A_2 v_2 = \lambda_2 v_2.$$

We then have v_2 perpendicular to v_1 and $0 < |\lambda_2| \leq |\lambda_1|$.
Taking further subsequences of H_1 yields by induction a decreasing sequence of subspaces

$$H_1 = V_1 \supset V_2 \supset V_3 \supset \cdots,$$

where $V_n = span\{v_1, \cdots, v_{n-1}\}^\perp$, restrictions $A_n = A|_{V_n}$ belongs to the compact self-adjoint operators in V_n and a sequence of real eigenvalues $\{\lambda_n\}_{n=1}^\infty$ with $|\lambda_n| = \|A_n\|$.

Then we can write

$$|\lambda_1| \geq |\lambda_2| \geq \cdots$$

with corresponding orthonormal eigensystem $\{\lambda_n, v_n\}_{n=1}^{\infty}$ of A. This procedure stops if the restriction of A_n to A to some H_{1_n} becomes zero. In this case, A has finitely many eigenvalues, then the sum is finite else infinite and A is said to be of finite rank if it has only finitely many eigenvalues. In the other case the non-zero eigenvalues cannot accumulate at some $\lambda \neq 0$. In other words, there does not exist a $\delta > 0$ such that $|\lambda_n| \geq \delta$ for all $n \in \mathbb{N}$; otherwise $Av_n = 0$ as $n \to \infty$. The sequence of orthogonal eigenvectors $\{v_n\}_{n=1}^{\infty}$ converges weakly to zero, and this yields a contradiction to the fact that non-zero eigenvalues cannot accumulate at some $\lambda \neq 0$ because

$$\|Av_n\| = |\lambda_n|\|v_n\| = |\lambda_n| \geq \delta > 0.$$

We now need to show the second part of the proof; that $\{v_n\}_{n=1}^{\infty}$ is an orthonormal basis of $Ker(A)^{\perp}$. The proof is trivial when A is degenerate, that it is a finite rank operator. In the non-degenerate case, let $x \in Ker(A)^{\perp}$, then we can consider

$$x' = \sum_{n=1}^{\infty} \langle x, v_n \rangle v_n \quad \in Ker(A)^{\perp}. \tag{28}$$

Since $(x - x')$ is perpendicular to v_n for all $n \in \mathbb{N}$, then it follows that

$$x - x' \in \bigcap_{n \in \mathbb{N}} H_{1_n},$$

and hence

$$\|A(x - x')\| = \|A_n(x - x')\| \leq \|A_n\|\|x - x'\| \leq |\lambda_n|\|x - x'\|$$

for all $n \in \mathbb{N}$. We know from above that $\lambda_n \to 0$ as $n \to \infty$, and it follows from this that $A(x - x') = 0$, that is, $x - x' \in Ker(A)$ and since $x - x' \in Ker(A)^{\perp}$, this shows that $x = x'$ and that $\{v_n\}_{n=1}^{\infty}$ is an orthonormal basis of $Ker(A)^{\perp}$. □

If A is compact linear but not self-adjoint, then A has no eigenvalues and therefore with no eigensystem. We can pass a substitute for an eigensystem $\{\lambda_n, v_n\}$ called the singular system $\{\sigma_n, u_n, v_n\}$ of the non-self-adjoint operator A. The construction is based on the connection between equation (1) and equation (18). Thus instead of A, one can consider A^*A. $A^*A : H_1 \to H_1$ is a compact linear self-adjoint operator.

Definition 2.20. *Let $A : H_1 \to H_2$, be a compact linear operator, a singular system*

$\{\sigma_n, u_n, v_n\}$ with $n \in \mathbb{N}$, $\sigma_n \geq 0$, $v_n \in H_1$ and $u_n \in H_2$ is called the singular system of A if the following conditions hold:

- *there exists a null sequence $\{\sigma_n\}$ with $\sigma_1 \geq \sigma_2 \geq \cdots > 0$ such that $\{\sigma_n^2\}$ is the sequence of the the non-zero eigenvalues of A^*A, which is also eigenvalues of AA^* ordered in non-increasing order and repeated according to the dimension of the associated eigenspace (that is with multiplicity);*
- *$\{v_n\} \subset H_1$ are the associated sequence of a complete orthonormal eigenvectors of A^*A which corresponds to $\{\sigma_n^2\}$ (spans $\overline{Ran(A^*)} = \overline{Ran(A^*A)}$);*
- *the sequence $\{u_n\} \subset H_2$ are the complete orthonormal system of eigenvectors of AA^* and span $\overline{Ran(A)} = \overline{Ran(AA^*)}$ and without loss of generality, $\{u_n\}$ are defined in terms of $\{u_n\}$ as*
$$u_n = \frac{Av_n}{\|Av_n\|}.$$

From the above construction, we have the following:

$$Av_n = \sigma_n u_n, \tag{29}$$
$$A^*u_n = \sigma_n v_n, \ \ \forall n \in \mathbb{N} \tag{30}$$

and we obtain using the singular system the following formulas

$$Ax = \sum_{n=1}^{\infty} \sigma_n \langle x, v_n \rangle u_n, \ \ \forall x \in H_1 \tag{31}$$
$$A^*y = \sum_{n=1}^{\infty} \sigma_n \langle y, u_n \rangle v_n, \ \ \forall y \in H_2. \tag{32}$$

The infinite sums in equation (31) and equation (32) converge in the corresponding norms of the Hilbert spaces H_1 and H_2 respectively. Thus, the sum converge due to the square integrability (existence of its norm) of the coefficients $\langle x, v_n \rangle$ and $\langle y, u_n \rangle$ respectively, the singular vectors' orthogonality and the boundedness of the singular values σ_n. Equations (31) and (32) are called the *singular value expansion of A* and are the analogous relation of the known *singular value decomposition* of an infinite dimensional case of a matrix.

The spectral theorem can be used to show the structure of the $Ran(A)$ and $Ran(A^*)$. This is described in the following theorem:

Theorem 2.21. *If H_1 and H_2 are separable Hilbert spaces and $A : H_1 \to H_2$ is a compact linear operator, then there exists complete orthonormal set of functions $\{v_n\} \subset H_1$ (right*

singular vectors), $\{u_n\} \subset H_2$ (left singular vectors), and a non-increasing numbers $\{\sigma_n\}$ such that

$$\begin{aligned} Av_n &= \sigma_n u_n, \\ A^* u_n &= \sigma_n v_n, \\ Ran(A^*)^\perp &= Ker(A), \\ Ran(A)^\perp &= Ker(A^*), \\ \overline{span\{v_n\}} &= \overline{Ran(A^*)} = Ker(A)^\perp, \\ \overline{span\{u_n\}} &= \overline{Ran(A)} = Ker(A^*)^\perp \end{aligned}$$

and the set $\{\sigma_n\}$ has non-zero limit points [31, 37].

Every non-zero singular value of a compact operator have finite multiplicity. For A to have only finitely many singular values, it is a necessary and sufficient condition for $Ran(A)$ to be finite-dimensional so that the singular values of all the infinite series in equation (31) and equation (32) degenerate into finite sums.

Example 2.22. *Suppose A is a linear Fredholm integral operator of the first kind such that*

$$\begin{aligned} A : L^2(\Omega) &\longrightarrow L^2(\Omega) \\ x &\longmapsto (Ax)(s) = \int_\Omega k(s,t)x(t)dt. \end{aligned}$$

The singular value expansion theorem states that the kernel k is an L^2-kernel and A is considered an operator on L^2 if and only if the kernel k is degenerate and is of the form

$$k(s,t) = \sum_{i=1}^{n} \phi_i(s)\psi_i(t), \quad s,t \in \Omega \tag{33}$$

with $n \in \mathbb{N}$, ϕ_i, $\psi_i \in L^2(\Omega)$. Above, n is the rank of the kernel.
At most there are countably many singular values with no accumulation other than 0. That is to say, we have

$$\lim_{n\to\infty} \sigma_n = 0.$$

If the operator A with closed range, $Ran(A)$ is finite-dimensional, then A has countably many singular values, $Ran(A)$ is not closed. This is seen as follows:
If $Ran(A)$ is a closed subspace of a Banach space then it is complete and due to the Open

Mapping Theorem the mapping

$$A|_{Ker(A)^{\perp}} : Ker(A)^{\perp} \to Ran(A)$$

is continuously invertible. It follows that

$$A(A|_{Ker(A)^{\perp}})^{-1} = I_{Ran(A)}$$

where $I_{Ran(A)}$, the identity operator is compact and thus, dim $Ran(A) < \infty$.

As a consequence of Theorem 2.12, we have the following proposition.

Proposition 2.23. *Let $A : H_1 \to H_2$ be a compact operator with dim $Ran(A) = \infty$, then the Moore-Penrose generalized inverse $A^{\dagger}$ is a closed, densely defined unbounded linear operator.*

Proof. The proof for Proposition 2.23 is given in [38]. □

Hence, from Proposition 2.23, if the compact linear operator has non-closed range then the best-approximate solution to equation (1) has no continuous dependence on the data that makes equation (1) ill posed. One can then use the singular system to create a representation of the Moore-Penrose generalized inverse $A^{\dagger}$. When there are only finitely many singular values, the infinite series in equations. (31) and (32) degenerate into finite sums. This is summarised in the following taking note from equation (19) in Theorem 2.9 and for $x^{\dagger} = A^{\dagger}y$; we have

$$\sum_{n=1}^{\infty} \sigma_n^2 \langle x^{\dagger}, v_n \rangle v_n = A^* A x^{\dagger} = A^* y = \sum_{n=1}^{n} \sigma_n \langle y, u_n \rangle v_n, \tag{34}$$

and by comparing the left and right sums of equation (34), we see that the respective linear component is obtained as

$$\langle x^{\dagger}, v_n \rangle = \frac{\langle y, u_n \rangle}{\sigma_n}.$$

A direct consequence of this is the following theorem which expresses the solution to equation (1) with a compact operator A in terms of a singular system.

Theorem 2.24. *Let $A : H_1 \to H_2$ be a compact linear operator and (σ_n, v_n, u_n), a singular system for A. The problem $Ax = y$ is solvable if and only if $y \in \mathcal{D}(A^{\dagger})$ and*

satisfies the Picard criterion given by

$$\|A^\dagger y\|^2 = \sum_{n=1}^{\infty} \frac{|\langle y, u_n\rangle|^2}{\sigma_n^2} < \infty. \tag{35}$$

In this case a best-approximate solution is given by the singular value expansion of $A^\dagger$ *and whenever* $y \in \mathcal{D}(A^\dagger)$

$$x^\dagger = A^\dagger y = \sum_{n=1}^{\infty} \frac{\langle y, u_n\rangle}{\sigma_n} v_n. \tag{36}$$

This representation of $A^\dagger y$ shows clearly that $A^\dagger$ is unbounded if $Ran(A)$ is infinite dimensional.

Proof. Let $y \in \mathcal{D}(A^\dagger)$, that is, $Qy = Ran(A) = Ker(A^*)^\perp$. The orthogonal projector Q onto $\overline{Ran(A)}$ can be written as

$$Q = \sum_{n=1}^{\infty} \langle \cdot, u_n\rangle v_n,$$

since the $\{u_n\}_{n=1}^{\infty}$ span $\overline{Ran(A)}$. Since $Qy = Ran(A)$, there exists an $x \in H_1$ with $Ax = Qy$; without loss of generality, we can assume that $x \in Ker(A)^\perp$. Since the sequence $\{v_n\}_{n=1}^{\infty}$ span $\overline{Ran(A^*)} = Ker(A)^\perp$, we can write $x = \sum_{n=1}^{\infty} \langle x, v_n\rangle v_n$, so that

$$\sum_{n=1}^{\infty} \langle y, u_n\rangle u_n = Ax = \sum_{n=1}^{\infty} \langle x, v_n\rangle A v_n = \sum_{n=1}^{\infty} \sigma_n \langle x, v_n\rangle u_n.$$

Thus, for all $n \in \mathbb{N}$, we have the identity

$$\langle y, u_n\rangle = \sigma_n \langle x, v_n\rangle. \tag{37}$$

Since, $(\langle x, v_n\rangle)_{n=1}^{\infty} \in l^2$, we have, by equation (37), that $\left(\frac{\langle y, u_n\rangle}{\sigma_n}\right)_{n=1}^{\infty} \in l^2$. This proves the condition in Picard's Criterion.
Conversely, assume that the Picard's condition holds. By the Riesz-Fischer Theorem from Functional analysis [39], we have

$$x := \sum_{n=1}^{\infty} \frac{\langle y, u_n\rangle}{\sigma_n} v_n \in H_1.$$

It follows that

$$Ax = \sum_{n=1}^{\infty} \frac{\langle y, u_n\rangle}{\sigma_n} A v_n = \sum_{n=1}^{\infty} \langle y, u_n\rangle u_n = Qy.$$

In particular, it holds that $Qy \in Ran(A)$ and hence $y \in \mathcal{D}(A^\dagger)$. Since the sequence $\{v_n\}_{n=1}^\infty$ span $Ker(A)^\perp$, we have $x \in Ker(A)^\perp$. Now,

$$\{z \in H_1 \mid Az = Qy\} = A^\dagger y + Ker(A). \tag{38}$$

Since x lies in both this set and in $Ker(A)^\perp$, it follows that x is the element of the minimal norm in this set, that is,

$$x = A^\dagger Qy = A^\dagger y.$$

Thus, equation (36) holds. This ends the proof. □

In Example 2.22, the properties of the kernel k determine the behaviour of σ_n and (v_n, u_n). The smoothness of the kernel determines how fast the σ_n decay to zero. The Picard's criterion says that a square integrable solution x of the equation $Ax = y$ can only exist if the absolute value of the coefficients $\langle y, u_n\rangle$ decay faster to zero than σ_n. It is seen from equation (36) that errors in the coefficients $\langle y, u_n\rangle$ which corresponds to singular functions with large n and small σ_n, i.e., high-frequency components are amplified by the factor σ_n^{-1} and are much dangerous than those for low-frequency components (large σ_n). In case of noisy data, if dim$(Ran(A)) = \infty$ and we perturb the right-hand side y in equation (1) by $y_n^\delta = y + \delta u_n$, then $\|y_n^\delta - y\| = \delta$, but

$$A^\dagger y - A^\dagger y_n^\delta = \sigma_n^{-1}\langle \delta u_n, u_n\rangle v_n$$

and hence

$$\|A^\dagger y - A^\dagger y_n^\delta\| = \delta\|A^\dagger u_n\| = \frac{\delta}{\sigma_n} \to \infty \text{ as } n \to \infty. \tag{39}$$

The amplification of the high-frequency error δ depends on the decay rate of σ_n. The faster the decay the higher the instability in Picard's criterion becomes (i.e., the higher the amplification of errors). This lack of stability in y is the basis for ill-posed problems posed in infinite dimensional function spaces and therefore the decay rate is used to classify the *degree of ill-posedness* of $Ax = y$ into three classes of ill-posedness. Hofmann [6, Definition 2.42], introduced the following definition:

We write $f(n) = \mathcal{O}(g(n))$, if there exists a constant $C > 0$ and $N > 0$ such that $f(n) \leq Cg(n)$ for all $n \geq N$.

Definition 2.25. *If there exists a real number $\alpha \in \mathbb{R}^+$ such that the singular values satisfy $\sigma_n = \mathcal{O}(n^{-\alpha})$ and this reads as "the singular values have an order of $n^{-\alpha}$ time complexity", then α is the degree of ill-posedness.*

1. *If* $0 < \alpha \leq 1$*, then we call the problem a mildly ill posed problem.*
2. *If* $\alpha > 1$*, then we call problem a moderately ill posed problem.*
3. *If* $\sigma_n = \mathcal{O}(e^{-\alpha n})$*, then the problem is called severely ill posed problem.*

The desire of developing methods capable of checking the instability in the Picard's criterion led Philips and Tikhonov in the early 1960s to introduce the principle of regularization for Fredholm integral equations. (see [40]).

3 REGULARIZATION THEORY

In the previous sections, we have seen that the major source of ill-posedness of inverse problems of the type in equation (1) is the fast decay of the singular values of A. In this section, the main objective is to compute an approximation to a solution of problem (1) from a noisy data y^δ with

$$\|y - y^\delta\| \leq \delta. \tag{40}$$

We shall introduce the notion of regularization and address other fundamental properties. In particular, a linear regularization method will be represented by a family of continuous operators $R_\alpha : H_2 \to H_1$, for $\alpha \in \mathbb{R}^+ \subset (0, \alpha_0)$, where the index set $\mathbb{R}^+$ includes at least one sequence $\alpha_n \to 0$. In fact, the regularization operator should converge to the generalized inverse in some sense as $\alpha \to 0$. This means that, as $\alpha \to 0$, we need the convergence $R_\alpha y \to A^\dagger y$ for $y \in \mathcal{D}(A^\dagger)$. For ill-posed operators, one typically has $Ran(A) \neq \overline{Ran(A)}$, that is the generalized inverse $A^\dagger$ is discontinuous and also if $y \in H_2 \setminus \mathcal{D}(A^\dagger)$, then $A^\dagger$ is unbounded and thus, we have to expect that $\|R_\alpha y\| \to \infty$ due to the unboundedness of $A^\dagger$ as seen in Theorem 2.14 and Proposition 2.23 in Section 2. Again, the unboundedness of $A^\dagger$ can be seen from equations (35) and (36), since for normalised u_n, we have equation (39). Therefore one have to use regularization methods in order to compute a stable approximation to the solution in the presence of noisy data.

3.1 Regularization of linear inverse problems

The theory of regularization for linear ill-posed problems is well developed in literature. For a good overview, see [2, 5, 9, 12].

In general, regularization is the approximation of the ill-posedness of an inverse problem by a solution of well-posed problems. We aim at approximating the best-approximate solution $x^\dagger = A^\dagger y$ of

$$Ax = y. \tag{41}$$

The exact information y is not known precisely, only y^δ with equation (40) is present. In literature, $y^\delta \in H_2$ is called the noisy data and $\delta > 0$ the noise level. According to Theorem 2.12, the generalised inverse $A^\dagger$ in general is not continuous, so in an ill-posed problem, $A^\dagger y^\delta$ is generally a bad approximation of $x^\dagger$ if it actually exists. Intuitively, regularizing equation (41) means essentially the construction of a family of continuous linear operators $\{R_\alpha\}$, by choosing a certain *regularization parameter* α in dependence

of δ, y^δ and A that approximate $A^\dagger$ (in some sense) and such that $x_\alpha^\delta := R_\alpha y^\delta$ satisfies the conditions above. A requirement for α for each $y \in \mathcal{D}(A^\dagger)$ is that, the regularized solution

$$R_\alpha y^\delta \longrightarrow A^\dagger y = x^\dagger \quad \text{as} \quad \delta \longrightarrow 0.$$

This desired convergence rate results is stated more precisely in the following equivalent definitions.

Definition 3.1. *Let $A : H_1 \to H_2$ be bounded operator between Hilbert spaces H_1 and H_2. Let $y \in Ran(A)$ and $y^\delta \in H_2$ with $\|y - y^\delta\| \leq \delta$.*
A family $\{R_\alpha\}$ of linear operators is called regularization (or regularization operator) of $A^\dagger$ if R_α are continuous linear operators between H_2 and H_1 (that is $R_\alpha : H_2 \to H_1$) and approximate $A^\dagger$ in the sense that

$$R_\alpha y \to A^\dagger y \quad \textit{as} \quad \alpha \to 0$$

for all $y \in \mathcal{D}(A^\dagger)$. Hence, a regularization is a point-wise approximation of the Moore-Penrose inverse with continuous operators.

The expectation is that α has to be chosen according to the error level δ. Let us consider the overall error $\|R_\alpha y^\delta - A^\dagger y\|$ that we split up into two parts by the triangle inequality as follows:

$$\begin{aligned} \|R_\alpha y^\delta - A^\dagger y\| &\leq \|R_\alpha y^\delta - R_\alpha y\| + \|R_\alpha y - A^\dagger y\| \\ &\leq \|R_\alpha\| \|y - y^\delta\| + \|R_\alpha y - A^\dagger y\| \end{aligned}$$

and thus

$$\|R_\alpha y^\delta - A^\dagger y\| \leq \delta \|R_\alpha\| + \|R_\alpha y - A^\dagger y\|. \tag{42}$$

We thus observe from equation (42) that the error between the exact and computed solutions consists of two parts; the first part describes data error multiplied by the condition number $\|R_\alpha\|$ of the regularized problem. The term $\|R_\alpha\| \to \infty$ as $\alpha \to 0$. The second part describes the approximation error which tends to zero with α, due to the point-wise convergence of R_α to $A^\dagger$.

Definition 3.2. *Let $A : H_1 \to H_2$ be a bounded linear operator between two Hilbert spaces H_1 and H_2 and let $\alpha_0 \in (0, +\infty]$. For every $\alpha \in (0, \alpha_0)$, let $R_\alpha : H_2 \to H_1$ be a family of continuous linear operators.*
The family $\{R_\alpha\}_{\alpha \in \mathbb{R}^+}$ is called a linear regularization operator for $A^\dagger$, if for all $y \in$

$\mathcal{D}(A^\dagger)$, there exists a function

$$\alpha : \mathbb{R}^+ \times H_2 \to (0, \alpha_0) \tag{43}$$

called the parameter choice rule for y, that allows to assign every $\alpha = \alpha(\delta, y^\delta)$ a corresponding operator $R_{\alpha(\delta,y^\delta)}$ and the regularized solution $x^\delta_{\alpha(\delta,y^\delta)} := R_{\alpha(\delta,y^\delta)} y^\delta$, and such that

$$\lim_{\delta\to 0} \sup\{\alpha(\delta, y^\delta) \mid y^\delta \in H_2, \|y - y^\delta\| \le \delta\} = 0. \tag{44}$$

If the parameter choice rule α also holds for

$$\lim_{\delta\to 0} \sup\{\|R_{\alpha(\delta,y^\delta)} y^\delta - A^\dagger y\| \mid y^\delta \in H_2, \|y - y^\delta\| \le \delta\} = 0, \tag{45}$$

then it is said to be convergent.
For a specific $y \in \mathcal{D}(A^\dagger)$, the pair (R_α, α) is called a (convergent) regularization method of equation (41) *if $\{R_\alpha\}$ is a regularization for $A^\dagger$ and α is a (convergent) parameter choice rule for y, that is equations* (44) *and* (45) *hold.*

Remark 3.3.

- *In the Definition 3.2, consideration is given to every feasible noisy data with noise level less than δ, the convergence is meant in the worst-case. Nevertheless, in a peculiar problem, a sequence of approximate solutions $x^{\delta_n}_{\alpha(\delta_n, y^\delta_n)}$ can converge fast to $x^\dagger$ also when equation* (45) *fails.*

- *The parameter choice rule $\alpha = \alpha(\delta, y^\delta)$ is dependent on the noise level and on the noisy data y^δ. From Definition 3.2, α depends on the exact solution y; which are not known in advance, so this becomes evident that the choice of α depends on δ, that is $\alpha = \alpha(\delta)$ is chosen a priori before the regularized solution is computed.*

Based on Remark 3.3, it is therefore advantageous to study strategies for the parameter choice rule α that depend on the numerical algorithm and are made during the algorithm (a posteriori). We state the three types of parameter choice rules in the following definition.

Definition 3.4. *Denote α as a parameter choice rule as in Definition 3.2. Then α is called*

1. *a-priori parameter choice rule if it does not depend on y^δ but on δ only. In this case, we write $\alpha = \alpha(\delta)$.*

2. *a-posteriori parameter choice rule if it depends on both δ and y^δ. In this case, we write $\alpha = \alpha(\delta, y^\delta)$.*

3. *heuristic parameter choice rule if it depends on y^δ only.*

Thus, if α is not dependent on y^δ, it can then be specified prior to real computations: this explains the a-priori and a-posteriori terms in the above definition.
In practice, one can be tempted to construct a parameter choice rule that is dependent only on the noisy data y^δ and independent of the noise level. However, the following popular results due to Bakushinskii [41] shows that such an approach cannot be convergent for an ill-posed problem.

Theorem 3.5 (Bakushinskii)**.** *Suppose $\{R_\alpha\}$ is a regularization for $A^\dagger$ such that for all $y \in \mathcal{D}(A^\dagger)$ there exists a convergent parameter choice rule α which is dependent on y^δ only. Then $A^\dagger$ is bounded.*

Proof. If $\alpha = \alpha(y^\delta)$, then it follows from equation (45) that for all $y \in \mathcal{D}(A^\dagger)$ we have

$$\lim_{\delta \to 0} \sup\{\|R_{\alpha(y^\delta)} y^\delta - A^\dagger y\| \mid y^\delta \in H_2, \|y - y^\delta\| \leq \delta\} = 0, \tag{46}$$

so that $R_{\alpha(y)} y = A^\dagger y$. Thus, by equation (46), if $\{y_n\}_{n \in \mathbb{N}}$ is a sequence in $\mathcal{D}(A^\dagger)$ converging to y, then

$$A^\dagger y_n = R_{\alpha(y_n)} y_n \to R_{\alpha(y)} y = A^\dagger y.$$

This implies that $A^\dagger$ is sequentially continuous on $\mathcal{D}(A^\dagger)$, hence it is bounded. □

Therefore, no heuristic parameter choice rule can produce a convergent method of regularization (R_α, α) in the ill-posed situation. This does not assume, though, that such a parameter choice rule cannot provide strong approximations of $x^\dagger$, given δ.

3.2 Regularization methods

In regularizing an ill-posed problem the obvious questions one needs to provide an answer to are as follows:

- How can I construct such regularization (regularization operators) as discussed earlier?
- How can I select a parameter choice rule to produce convergent methods of regularization?
- How can I perform these steps in some optimal way?

These questions will be dealt with in this and the following sections. The result below provides a description of regularization operators, thus answers the first question by this basic property; it can be shown that for any regularization an a-priori parameter choice rule, and thus, a convergent regularization, exists.

Proposition 3.6. *Let $\{R_\alpha\}_{\alpha>0}$ be a family of continuous operators. Then from Definition 3.1, there exists an a-priori parameter choice rule α, such that (R_α, α) is a convergent regularization for equation* (41).

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

$$\|R_{\sigma(\varepsilon)} y^\delta - A^\dagger y\| \leq \frac{\varepsilon}{2}.$$

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

$$\|R_{\sigma(\varepsilon)} z - R_{\sigma(\varepsilon)} y\| \leq \frac{\varepsilon}{2} \quad \text{for all } \ z \in H_2 \ \text{ if } \ \|z - y\| \leq \rho(\varepsilon).$$

Without loss of generality we can assume that ρ is continuous, strictly monotone increasing function with $\lim_{\varepsilon\to 0} \rho(\varepsilon) = 0$. Then there is a strictly monotone and continuous inverse function ρ^{-1} on the range $Ran(\rho)$ with $\lim_{\delta\to 0} \rho^{-1}(\delta) = 0$. We now continuously extend ρ^{-1} on $\mathbb{R}^+$ and define the a-priori parameter choice rule strategy as

$$\begin{aligned} \alpha : \mathbb{R}^+ &\to \mathbb{R}^+ \\ \delta &\to \sigma(\rho^{-1}(\delta)). \end{aligned}$$

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

$$\|R_{\alpha(\delta)} y^\delta - A^\dagger y\| \leq \|R_{\sigma(\varepsilon)} y^\delta - R_{\sigma(\varepsilon)} y\| + \|R_{\sigma(\varepsilon)} y^\delta - A^\dagger y\| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

follows for all $y^\delta \in H_2$ if $\|y - y^\delta\| \leq \delta$. Thus (R_α, α) is a convergent regularization method for equation (41) and the function α defines an a-priori parameter choice rule. $\square$

Remark 3.7. *Conversely from Proposition 3.6, if (R_α, α) is a convergent regularization method, then we can conclude from equation* (44) *that*

$$\lim_{\delta\to 0} R_{\alpha(\delta, y^\delta)} y = A^\dagger y, \quad \textit{with } \ y \in \mathcal{D}(A^\dagger),$$

and α is continuous with respect to δ, then this implies

$$\lim_{\sigma \to 0} R_\sigma y = A^\dagger y \quad \textit{as} \quad \sigma \to 0.$$

Therefore, the correct approach to understanding the concept of regularization is pointwise convergence of the regularization operators. Furthermore, in the case of $y \notin \mathcal{D}(A^\dagger)$, as the generalised inverse is not defined for those functions, we cannot expect that R_α, a convergent regularization remain bounded as $\alpha \to 0$, since then $A^\dagger$ would have to be bounded. This is confirmed by the following result.

Proposition 3.8. *Let $\{R_\alpha\}_{\alpha>0}$ be a continuous linear regularization of $A^\dagger$. Let*

$$x_\alpha := R_\alpha y. \tag{47}$$

as defined in Definition 3.1. Moreover, if

$$\sup_{\alpha>0} \|AR_\alpha\| < \infty, \tag{48}$$

then

$$\|x_\alpha\| \to \infty \ \textit{for} \ y \notin \mathcal{D}(A^\dagger). \tag{49}$$

Proof. In the case for $y \in \mathcal{D}(A^\dagger)$ in equation (47), the convergence of x_α is centred on Proposition 3.6 above. We then only need to look at the case when $y \notin \mathcal{D}(A^\dagger)$. Now, assume that there is a sequence $\alpha_n \to 0$ such that $\|x_{\alpha_n}\|$ is uniformly bounded. Then there is a weakly convergent subsequence x_{α_n} with some limit $x \in H_1$. As continuous linear operators are also weakly continuous, we have $Ax_{\alpha_n} \to Ax$. However, as AR_α are uniformly bounded operators, we also conclude $Ax_{\alpha_n} = AR_{\alpha_n} y \to Qy$. Hence, $Ax = Qy$ and consequently $y \in \mathcal{D}(A^\dagger)$ is a contradiction to the assumption $y \notin \mathcal{D}(A^\dagger)$. In conclusion, for $y \notin \mathcal{D}(A^\dagger)$, no bounded sequence $\|x_{\alpha_n}\|$ can exist, hence equation (49) holds. □

We finally can characterise an a-priori parameter choice rule α that lead to convergent regularization methods by the following Proposition.

Proposition 3.9. *Let $\{R_\alpha\}_{\alpha>0}$ be a linear regularization, and $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ an a-priori parameter choice rule. Then (R_α, α) is a convergent regularization method if and only if*

$$\lim_{\delta \to 0} \alpha(\delta) = 0 \tag{50}$$

and

$$\lim_{\delta \to 0} \delta \|R_{\alpha(\delta)}\| = 0 \tag{51}$$

hold.

Proof. If equations (50) and (51) hold, then for every $y^\delta \in H_2$ with $\|y^\delta - y\| \leq \delta$, we have

$$\begin{aligned} \|R_{\alpha(\delta)} y^\delta - A^\dagger y\| &\leq \|x_{\alpha(\delta)} - A^\dagger y\| + \|x_{\alpha(\delta)} - R_{\alpha(\delta)} y^\delta\| \\ &\leq \|x_{\alpha(\delta)} - A^\dagger y\| + \delta \|R_{\alpha(\delta)}\|. \end{aligned}$$

Due to equations (47), (50) and (51), the right-hand side of the inequality converges to zero and thus (R_α, α) is a convergent regularization method. We now show the converse. Now let (R_α, α) be a convergent regularization method and assume that equation (51) does not hold, so that there exists a sequence $\delta_n \to 0$ such that $\|\delta_n R_{\alpha(\delta_n)}\| \geq C > 0$ for some constant C. Then we can find a sequence $\{z_n\}$ in H_2 with $\|z_n\| = 1$ such that $\delta_n \|R_{\alpha(\delta_n)} z_n\| \geq \frac{C}{2}$. Then for any $y \in \mathcal{D}(A^\dagger)$ and $y_n := y + \delta_n z_n$, we obtain $\|y - y_n\| \leq \delta_n$, but

$$R_{\alpha(\delta_n)} y_n - A^\dagger y = (R_{\alpha(\delta_n)} y - A^\dagger y) + \delta_n R_{\alpha(\delta_n)} z_n$$

does not converge to 0, since the second term $\delta_n R_{\alpha(\delta_n)} z_n$ is unbounded. Hence, for sufficiently small δ_n, equation (45) is violated for $y^\delta = z_n$ and thus, (R_α, α) is not a convergent regularization method. □

Now we consider an example of a regularization constructed to fit the definitions above. Refer to [8] for more examples.

Example 3.10. *Consider the operator $A : L^2[0,1] \to L^2[0,1]$ defined by*

$$(Ax)(s) := \int_0^s x(t)dt.$$

Then A is a bounded linear compact operator and it is easily seen that

$$Ran(A) = \{y \in \mathcal{W}^{2,1}[0,1] \mid y \in \mathcal{C}([0,1]), y(0) = 0\} \tag{52}$$

where $\mathcal{W}^{2,1}$ denote the Sobolev space on $[0,1]$ with order 1 for the L^2 space and $\mathcal{C}$ is the set of continuous functions on $[0,1]$. The distributional derivative from $Ran(A)$ to $L^2[0,1]$ is the inverse of A. Since $\mathcal{C}_0^\infty[0,1] \supseteq Ran(A)$, $Ran(A)$ is dense in $L^2[0,1]$ and

$Ran(A)^{\perp} = \{0\}$. *For* $y \in \mathcal{C}([0,1])$ *and* $\alpha > 0$, *define*

$$(R_\alpha y)(t) := \begin{cases} \frac{1}{\alpha}(y(t+\alpha) - y(t)), & \textit{if} \;\; 0 \le t \le \frac{1}{2}, \\ \frac{1}{\alpha}(y(t) - y(t-\alpha)), & \textit{if} \;\; \frac{1}{2} < t \le 1. \end{cases} \tag{53}$$

Then $\{R_\alpha\}$ *is a family of linear and bounded operators with*

$$\|R_\alpha y\|_{L^2[0,1]} \le \frac{\sqrt{6}}{\alpha}\|y\|_{L^2[0,1]} \tag{54}$$

defined on a dense linear subspace of $L^2[0,1]$, *thus it an be extended to the whole* $L^2[0,1]$ *and equation* (54) *holds. Since the measure of* $[0,1]$ *is finite, for* $y \in Ran(A)$ *the distributional derivative of* y *lies in* $L^1[0,1]$, *so* y *is a function of bounded variation. By the Lebesgue's Theorem, the derivative* y' *exists almost everywhere in* $[0,1]$ *and it is equivalent to the distributional derivative of* y *as an* L^2-*function. We can therefore use the Dominate Convergence Theorem to prove that*

$$\|R_\alpha y - A^\dagger y\|_{L^2[0,1]} \to 0, \;\; \textit{as} \;\; \alpha \to 0$$

so that, according to Proposition 3.6, R_α *is a regularization for the distributional derivative* $A^\dagger$.

4 LANDWEBER ITERATION

4.1 Introduction

Let us consider equation (1) and that equation (40) is satisfied given a perturbed data y^δ. The idea of most iterative methods is to approximate $A^\dagger y$ with a sequence of iterates $\{x_k\}_{k\in\mathbb{N}}$ and are based on the transformation of the normal equation (18) into equivalent fixed point equations such as

$$x = x + A^*(y - Ax) = (I - A^*A)x + A^*y \tag{55}$$

[2, 20, 42]. The vector $A^*(y - Ax)$ is the directional negative gradient of the quadratic functional

$$x \longmapsto \|Ax - y\|^2.$$

Landweber [43] in 1951 established a very strong convergence provided A is compact and $y \in \mathcal{D}(A^\dagger)$, Fridman [44] studied other properties of A; not only compact, but also being a self-adjoint positive semi-definite operator. Bialy [45] on the other hand in 1959 extended the results of Landweber and Fridman to not necessarily a compact operator A. The Landweber iteration is given an appropriate initial guess say x^* which selects the particular solution which will be approximated in case one is given a noisy data y^δ instead of y and using $x_0^\delta = x^*$ the iteration computes the sequence of iterates $\{x_k^\delta\}_{k\in\mathbb{N}}$ recursively.
The Landweber iteration is defined as follows:

Definition 4.1 (Landweber Iteration). *Fix any appropriate initial guess $x_0^\delta = x^* \in H_1$ and for $k = 1, 2, \ldots$ compute the Landweber approximations recursively using the formula*

$$x_k^\delta = x_{k-1}^\delta + A^*(y^\delta - Ax_{k-1}^\delta). \tag{56}$$

As observed in the Definition 4.1, one can conveniently assume without loss of generality

$$\|A\| \leq 1, \tag{57}$$

in which case $I - A^*A$ and $I - AA^*$ are both positive semi-definite operators with at most norm one as seen in [44], since

$$\|T\| = \sup_{\|x\|_{H_1}=1} |\langle x, Tx\rangle_{H_1}|$$

for any self-adjoint operator $T : H_1 \to H_1$ and $T = A^*A = AA^*$ [46]. If it was not the case as in equation (57), then one would introduce a fixed relaxation parameter $\tau > 0 \in \mathbb{R}$ with $0 < \tau \leq \dfrac{1}{\|A\|^2}$ that precedes A^*, in equation (56). That is, the iteration would be

$$x_k^\delta = x_{k-1}^\delta + \tau A^*(y^\delta - Ax_{k-1}^\delta) = (I - \tau A^*A)x_{k-1}^\delta + \tau A^* y^\delta, \;\; k \in \mathbb{N}. \tag{58}$$

This iteration scheme is a special case of the steepest descent algorithm applied to the quadratic functional $\|Ax - y\|_2^2$ and is seen in the following lemma

Lemma 4.2. *Let the sequence $\{x_k^\delta\}$ be defined by equation* (58) *and define the quadratic functional $\Psi : H_1 \to \mathbb{R}$ by $\Psi(x) = \frac{1}{2}\|Ax - y^\delta\|_2^2$. Then Ψ is Fréchet differentiable in each $z \in H_1$ and*

$$\Psi'(z)x = (Az - y^\delta, Ax) = (A^*(Az - y^\delta), x), \;\; x \in H_1. \tag{59}$$

The linear functional $\Psi'(z)$ can be identified with $A^*(Az - y^\delta) \in H_1$ in the Hilbert space H_1 over the field of real numbers $\mathbb{R}$.

Therefore $x_k^\delta = x_{k-1}^\delta + \tau A^*(y^\delta - Ax_{k-1}^\delta)$ is the steepest descent step with step-size τ.

Equivalently, one could multiply the equation $Ax = y^\delta$ by $\sqrt{\tau}$ and perform iteration with equation (56).
Furthermore, following from equation (58), if $\{z_k^\delta\}$ is the sequence of iterates with initial guess value $z_0^\delta = 0$ and the data $y^\delta - Ax_0^\delta$, then $x_k^\delta = x_0^\delta + z_k^\delta$. So one can assume without loss of generality that the standard choice of initial guess is that $x_0^\delta = 0$.
If $\|A\|^2 = 1 < 2$ then the associated fixed point $I - A^*A$ in equation (55) is nonexpansive and the method of successive approximations may be applied [2, 42]. For ill-posed problems, the fixed point operator $I - A^*A$ is no contraction. This is because the spectrum of A^*A clusters at the starting point (origin). For instance, if A is compact, then there exists a set $\{\lambda_n\}$ of eigenvalues of A^*A such that $|\lambda_n| \to 0$ as $n \to \infty$ and with its associated eigenvectors $\{v_n\}$ we have

$$\|(I - A^*A)v_n\|\|v_n\|^{-1} = \|(1 - \lambda_n)v_n\|\|v_n\|^{-1} = |1 - \lambda_n| \longrightarrow 1 \;\text{ as }\; n \to \infty.$$

That is $\|I - A^*A\| \leq 1$.
The following theorem is the work of Landweber [43] where he proved the strong convergence of compact operators.

Theorem 4.3. *If $y \in \mathcal{D}(A^\dagger)$, then the Landweber approximations x_k corresponding to the true data y converge to $A^\dagger y$, i.e., $x_k \to A^\dagger y = x^\dagger$ as $k \to \infty$. If $y \notin \mathcal{D}(A^\dagger)$, then*

$\|x_k\| \to \infty$ *as* $k \to \infty$.

Proof. By mathematical induction, the iteration terms x_k may be written non-recursively in the form

$$x_k = \sum_{j=0}^{k-1} (I - A^*A)^j A^* y. \tag{60}$$

Next if we suppose that $y \in \mathcal{D}(A^\dagger)$, then for $A^*y = A^*Ax^\dagger$, we have

$$x^\dagger - x_k = x^\dagger - \sum_{j=0}^{k-1} (I - A^*A)^j A^*Ax^\dagger = x^\dagger - A^*A \sum_{j=0}^{k-1} (I - A^*A)^j x^\dagger$$

and

$$A^*A \sum_{j=0}^{k-1} (I - A^*A)^j = I - (I - A^*A)^k \tag{61}$$

thus

$$x^\dagger - x_k = x^\dagger - (I - (I - A^*A)^k)x^\dagger = (I - A^*A)^k x^\dagger.$$

We can denote functions g and r as

$$g_k(\lambda) = \sum_{j=0}^{k-1} (1 - \lambda)^j \quad \text{and} \quad r_k(\lambda) = (1 - \lambda)^k, \tag{62}$$

where $g_k(\lambda)$ and $r_k(\lambda)$ are parameter-dependent family of functions which are piecewise continuous on $[0, \|A\|^2]$ (that is, on a set containing the spectrum A^*A). Since $\|A\| \leq 1$ as we have assumed before, we consider $\lambda \in (0, 1]$: in this interval $\lambda g_k(\lambda) = 1 - r_k(\lambda)$ is uniformly bounded and $g_k(\lambda)$ converges to $\frac{1}{\lambda}$ as $k \to \infty$ because $r_k(\lambda)$ converges to 0. □

Theorem 4.3 states that the approximation error of the Landweber iteration converges to 0 when $y \in \mathcal{D}(A^\dagger)$. However, what happens if we have a perturbed data y^δ? We must then examine the behaviour of the propagated data error. According to Theorem 4.3, for a noisy data y^δ with $y^\delta \notin \mathcal{D}(A^\dagger)$ the iterates must diverge; on the other hand the $k - th$ iterate x_k^δ for fixed k is continuously dependent on the data. This is seen in the following theorem and result.

Theorem 4.4. *For fixed iteration index k, the iterate x_k^δ depends continuously on the perturbed data y^δ, since x_k^δ is the result of a combination of continuous operations and*

for $k \in \mathbb{N}$, the linear and bounded operator $R_k : H_2 \to H_1$ defined by

$$x_k^\delta = R_k y^\delta \quad \textit{with} \quad R_k = \sum_{j=0}^{k-1} (I - A^*A)^j A^* \tag{63}$$

holds. Assuming further that $\|A\| \leq 1$ and that A is injective with dense range $Ran(A) \subset H_2$, if $y^\delta \in Ran(A)$, then $x_k^\delta \to A^{-1}y^\delta$ as $k \to \infty$, and if $y^\delta \notin Ran(A)$, then $\|x_k^\delta\| \to \infty$ as $k \to \infty$.

Proof. It is obvious to see that $R_k : y^\delta \mapsto x_k^\delta$ is continuous, because x_k^δ is the result of a (finite) combination of continuous transformations of the given data. The particular form of R_k is readily obtained by mathematical induction: Since $x_0^\delta = 0$, we have $x_1^\delta = A^*y^\delta$, and hence $R_1 = A^*$ as claimed; for $k > 1$ it follows from equation (56) and the induction hypothesis that

$$x_k^\delta = (I - A^*A)R_{k-1}y^\delta + A^*y^\delta = \left(A^* + \sum_{j=0}^{k-2} (I - A^*A)^{j+1} A^* \right) y^\delta,$$

and this coincides with equation (63).
For $y^\delta \in Ran(A)$ we have $y^\delta = Ax^\delta$ for some $x^\delta \in H_1$, and it follows from equation (56) that

$$x^\delta - x_k^\delta = x^\delta - x_{k-1}^\delta - A^*A(x^\delta - x_{k-1}^\delta) = (I - A^*A)(x^\delta - x_{k-1}^\delta).$$

And by induction, this implies that

$$x^\delta - x_k^\delta = (I - A^*A)^k x^\delta, \quad k \in \mathbb{N}. \tag{64}$$

If $\|A\| \leq 1$, then the norm of the operators $(I - A^*A)^k$ are uniformly bounded by 1, and for each $x^\delta \in Ran(A^*)$, that is for $x^\delta = A^*w$, we have

$$(I - A^*A)^k x^\delta = (I - A^*A)^k A^* w \longrightarrow 0, \quad as \quad k \to \infty.$$

Accordingly, by the assumption that A is injective with dense range $Ran(A) \subset H_2$ implies that $Ran(A^*)$ is dense in H_1, and that $(I - A^*A)^k$ converges pointwise to 0 on a dense subset of H_1, and by the Banach-Steinhaus theorem thus the iteration error in equation (64) converges to 0 for every $x^\delta \in H_1$ as $k \to \infty$.
From equation (63) it follows that

$$AR_k = A \sum_{j=0}^{k-1} (I - A^*A)^j A^* = AA^* \sum_{j=0}^{k-1} (I - AA^*)^j = I - (I - AA^*)^k.$$

Now, we have

$$\|AR_k\| \leq 1, \tag{65}$$

and hence if A is injective with dense range in H_2, and if $y^\delta \notin Ran(A)$, then the iterates diverge to ∞ in the norm as $k \to \infty$. □

Thus, for fixed k, x_k^δ is continuously dependent on the data so that the error of propagation cannot be arbitrarily high. The following result is as a consequence of this.
This results in the following.

Proposition 4.5. *Let y, $y^\delta \in H_2$ be a pair of data with equation* (40) *and let $\{x_k\}_{k=1}^\infty$ and $\{x_k^\delta\}_{k=1}^\infty$ be their respective iteration sequences. We then have*

$$\|x_k - x_k^\delta\| \leq \sqrt{k}\delta, \quad k \geq 0. \tag{66}$$

Proof. By equation (60), we have

$$\begin{aligned} x_k - x_k^\delta &= \sum_{j=0}^{k-1}(I - A^*A)^j A^*(y - y^\delta) \\ &= R_k(y - y^\delta), \end{aligned}$$

and we are to find the norm of R_k. From equation (61) follows

$$\|R_k\|^2 = \|R_k R_k^*\| = \|\sum_{j=0}^{k-1}(I - A^*A)^j(I - (I - A^*A)^k)\| \leq \|\sum_{j=0}^{k-1}(I - A^*A)^j\|,$$

where $I - AA^*$ is positive semi-definite with $\|I - AA^*\| \leq 1$. It is obvious that the right hand side of the inequality is bounded by k, and the assertion follows. □

Remark 4.6. *From previous theorems and results, there are two components of the the total error, as seen in equation* (42)*; an approximation error with slow convergence to* 0 *and the data error propagation of the order of at most $\sqrt{k}\delta$. From equation* (42)*, the data error is multiplied by the condition number $\|R_\alpha\|$ and by comparing this with equation* (66)*, then $\|R_\alpha\| = \sqrt{k}$. If k is small, then the computed approximation x_k is an over smoothed solution [47]. That is the data error in equation* (42) *is negligible (the difference between x_k and x_k^δ in equation* (66) *is very small) and the total error converge to the true solution $A^\dagger y$, but when $\sqrt{k}\delta$ approaches the order of the degree of the approximation error, the data error is now seen in x_k^δ and the total error starts to increase until there is worst-case rate of divergence.*

Remark 4.7. *The phenomenon observed in Remark 4.6 has been termed semi-convergence behaviour of iterative methods by Natterer [48], see also [49,50]. The regularizing effects of iterative techniques is efficient if one finds a realistic stopping rule or criteria to detect the transient between convergence and divergence. This means that the iteration index k takes the role of the regularization parameter α and the stopping rule works similarly as the parameter choice rule for continuous regularization methods [51]. That is, in the case of a noisy data, the iteration procedure is combined with a stopping rule in order to serve as a regularization method and one should terminate the iteration procedure by an appropriate stopping rule that involve the noise and noise level δ. This means the iteration in equation* (58) *is stopped after $k_* = k_*(\delta, y^\delta)$, where k_* is the stopping index [42, 52]. A generalized principle employed here is the most well-known stopping rule called the discrepancy principle of Morozov [53] which we will discuss fully in the next subsection.*

4.2 Connection of the Singular Value Expansion and the Landweber's Iteration

In this section, we discuss the usefulness of the singular value expansion in the convergence of the Landweber iteration when considering both an exact and noisy data.

4.2.1 Exact Data

By using the definition of the Landweber iteration as defined in Definition 4.1 with equation (58) but without the presence of noise level δ, that is when dealing with exact data, one has

$$x_k = x_{k-1} + \tau A^*(y - Ax_{k-1}) = (I - \tau A^* A)x_{k-1} + \tau A^* y, \tag{67}$$

for some $\tau > 0$ and $x_0 \equiv 0$. Let us initially assume that $y \in \mathcal{D}(A^\dagger)$ and with the singular value expansion of A and A^* we obtain an equivalent form of equation (67) as

$$\sum_{n=1}^{\infty} \langle x_k, v_n \rangle v_n = \sum_{n=1}^{\infty} (1 - \tau\sigma_n^2)\langle x_{k-1}, v_n \rangle v_n + \tau\sigma_n \langle y, u_n \rangle v_n. \tag{68}$$

and since vectors v_n are orthogonal we have,

$$\langle x_k, v_n \rangle = (1 - \tau\sigma_n^2)\langle x_{k-1}, v_n \rangle + \tau\sigma_n \langle y, u_n \rangle \tag{69}$$

for any $n \in \mathbb{N}$. Assuming $x_0 \equiv 0$, summing equation (69) yields

$$\langle x_{k-1}, v_n \rangle = \tau\sigma_n \langle y, u_n \rangle \sum_{i=1}^{k-1} (1 - \tau\sigma_n^2)^{(k-1)-i}. \tag{70}$$

We can derive

$$\sum_{i=1}^{k-1} (1 - \tau\sigma_n^2)^{(k-1)-i} = \frac{1 - (1 - \tau\sigma_n^2)^{k-1}}{\tau\sigma_n^2}. \tag{71}$$

Therefore, by inserting equation (71) into (70) we obtain

$$\begin{aligned} \langle x_{k-1}, v_n \rangle &= \tau\sigma_n \langle y, u_n \rangle \left(\frac{1 - (1 - \tau\sigma_n^2)^{k-1}}{\tau\sigma_n^2} \right) && (72) \\ &= \left(1 - (1 - \tau\sigma_n^2)^{k-1}\right) \frac{1}{\sigma_n} \langle y, u_n \rangle. && (73) \end{aligned}$$

The direct consequence of equation (73) is seen as

$$\langle x_{k-1}, v_n \rangle \to \langle x^\dagger, v_n \rangle = \frac{1}{\sigma_n} \langle y, u_n \rangle$$

if we ensure $(1-\tau\sigma_n^2)^{k-1} \to 0$. That is to say, we have to choose τ such that $|1-\tau\sigma_n^2| < 1$ (so respectively $0 < \tau\sigma_n^2 < 2$) for all n in order to obtain convergence of the singular value expansion. We exploit that $\sigma_1 = \|A\| > \sigma_n$ for all n and select τ such that

$$0 < \tau < \frac{2}{\|A\|^2}$$

is satisfied. When we view the iteration number as the regularization parameter $\alpha := \frac{1}{k}$, then we obtain the regularization method

$$x_\alpha = R_\alpha y = \sum_{n=1}^{\infty} g_\alpha(\sigma_n) \langle y, u_n \rangle$$

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

4.2.2 Noisy Data

We employ the same iteration procedure as in the case of the exact data but with the presence of a noise level δ. We do this to explain the iterates' dependency on the noise.

We use equation (58) in this case. It is given as:

$$x_k^\delta = x_{k-1}^\delta + \tau A^*(y^\delta - Ax_{k-1}^\delta) = (I - \tau A^* A)x_{k-1}^\delta + \tau A^* y^\delta, \;\; k \in \mathbb{N}. \tag{74}$$

Utilizing the singular values of A and A^*, one can write out the error as

$$\begin{aligned} \langle x_{k-1}^\delta - x^\dagger, v_n\rangle &= \left(1 - (1-\tau\sigma_n^2)^{k-1}\right)\frac{1}{\sigma_n}\langle y^\delta, u_n\rangle - \frac{1}{\sigma_n}\langle y, u_n\rangle & (75)\\ &= \frac{\left(1 - (1-\tau\sigma_n^2)^{k-1}\right)}{\sigma_n}\langle y^\delta - y, u_n\rangle + \frac{(1-\tau\sigma_n^2)^{k-1}}{\sigma_n}\langle y, u_n\rangle. & (76) \end{aligned}$$

Since the second term in the sum in equation (76) decays exponentially to zero as $k \to \infty$,

$$\frac{(1-\tau\sigma_n^2)^{k-1}}{\sigma_n}\langle y, u_n\rangle = (1-\tau\sigma_n^2)^{k-1}\langle x^\dagger, u_n\rangle.$$

Using equation (71), the absolute value of the first term can be estimated by

$$\frac{\left(1 - (1-\tau\sigma_n^2)^{k-1}\right)}{\sigma_n}|\langle y^\delta - y, u_n\rangle| = \tau\sigma_n \sum_{j=0}^{k-2}(1-\tau\sigma_n^2)^j|y^\delta - y, u_n| \le \tau\sigma_n k\delta$$

if we take for any $k > 1$. One can end the iteration procedure by a stopping rule k_* and this rule involves the noisy data and the noise level (see Remark 4.7). Thus, the iteration in Equation (74) can only be terminated after $k_* = k_*(\delta, y^\delta)$ and the solution obtained is known as the regularized solution which is then $x_{k_*}^\delta$. We can estimate the components of the error between the exact and regularized solutions by employing the singular value expansion with the above arguments. This is given as follows

$$|\langle x_{k_*(\delta)}^\delta - x^\dagger, v_n\rangle| \le \tau\sigma_n k_*\delta + (1-\tau\sigma_n^2)^{k-1}\|x^\dagger\|.$$

The Landweber iteration is a convergent regularization in the sense that we choose the stopping index in such a way that $k_*(\delta) \to \infty$ and $k_\delta(\delta) \to 0$ as $\delta \to 0$, then all components in the sum converge to zero and hence, $x_{k_*(\delta)}^\delta \to x^\dagger$.

This leads us to the discussion of such stopping rules and the analysis of the Discrepancy Principle which describes the above section in full details in the next section.

4.3 The Discrepancy Principle

Unlike the a-priori choice of the regularization parameter α, one may try to find strategies for selecting α where results occurring during the computations are used. Such strategies are called a-posteriori strategies. For the Landweber iteration, it is easier to consider an a-posteriori parameter stopping rule such as the generalized discrete version of the *Discrepancy Principle*.

Definition 4.8. *For $k = 0, 1, 2, \cdots$ let x_k^δ be the kth iterate of an iterative method to solve $Ax = y$ with noisy data y^δ such that $\|y - y^\delta\| \leq \delta$, $\delta > 0$. The stopping index $k_* = k_*(\delta, y^\delta)$ which corresponds to the discrepancy principle is the smallest k such that*

$$\|y^\delta - Ax_k^\delta\| \leq \eta\delta \tag{77}$$

with fixed number $\eta > 1$.

Vainikko [54] successfully applied this to the regularization of linear ill-posed problems by Landweber iteration. Defrise and De Mol [55] combined the Landweber iteration with the discrepancy stopping rule: stop when

$$\|y^\delta - Ax_k^\delta\| \leq \frac{2}{2 - \tau\|A\|^2}\delta$$

is satisfied for the first time, and take $x_{k_*}^\delta$ as an approximation to a solution x_* of $Ax = y$ [56].

It has to be shown that this stopping index k_* is well-defined, that is, there is a k finite index so that the residual discrepancy $\|y^\delta - Ax_k^\delta\|$ is smaller than the tolerance $\eta\delta$. The residual of the Landweber iteration can be written in the form

$$y^\delta - Ax_k^\delta = y^\delta - Ax_{k-1}^\delta - AA^*(y^\delta - Ax_{k-1}^\delta) = (I - AA^*)(y^\delta - Ax_{k-1}^\delta), \tag{78}$$

To implement the stopping rule, it is necessary to monitor this residual $y^\delta - Ax_k^\delta$ and compare its norm with the level of noise. Hence, from the non-expansivity of $I - AA^*$ follows that the residual norm decreases monotonically. However, the monotonicity is not a guarantee that the discrepancy principle is well defined and definitely one needs to show a more precise estimate of the residual norm. This leads us to the following proposition as presented in [2].

Proposition 4.9. *Let $y \in Ran(A)$ and consider any solution x of the equation $Ax = y$.*

A sufficient condition for x_{k+1}^δ to be a better approximation of $x^\dagger$ than x_k^δ is that

$$\|y^\delta - Ax_k^\delta\| > 2\delta. \tag{79}$$

Proof. To understand how the discrepancy principle works, let us consider the error during the iteration by estimating

$$\begin{aligned}
\|x^\dagger - x_{k+1}^\delta\|^2 &= \|x^\dagger - x_k^\delta - A^*(y^\delta - Ax_k^\delta)\|^2 \\
&= \|x^\dagger - x_k^\delta\|^2 - 2\langle x^\dagger - x_k^\delta, A^*(y^\delta - Ax_k^\delta)\rangle + \langle y^\delta - Ax_k^\delta, AA^*(y^\delta - Ax_k^\delta)\rangle \\
&= \|x^\dagger - x_k^\delta\|^2 - 2\langle y - y^\delta, y^\delta - Ax_k^\delta\rangle - \|y^\delta - Ax_k^\delta\|^2 \\
&+ \langle y^\delta - Ax_k^\delta, (AA^* - I)(y^\delta - Ax_k^\delta)\rangle.
\end{aligned}$$

As $AA^* - I$ is negative semidefinite, that is $AA^* - I \leq 0$ we obtain

$$\|x^\dagger - x_k^\delta\|^2 - \|x^\dagger - x_{k+1}^\delta\|^2 \geq \|y^\delta - Ax_k^\delta\|(\|y^\delta - Ax_k^\delta\| - 2\delta). \tag{80}$$

Thus since $k < k_*$ and if $\|y^\delta - Ax_k^\delta\| > 2\delta$, then one can guarantee that the right hand side of equation (80) is positive and hence

$$\|x^\dagger - x_{k+1}^\delta\|^2 < \|x^\dagger - x_k^\delta\|^2,$$

that is the error decreases until the iteration is stopped. □

This is a good motivation for the application of the discrepancy principle as an a-posteriori stopping criterion when carry out an iterative method such as the Landweber iteration since it has proven to be a convergent regularization method. Proposition 4.9 leads to the following results.

Proposition 4.10. *For fixed $\eta > 1$ in equation* (77)*, the Discrepancy Principle determines a finite stopping index $k_*(\delta, y^\delta)$ for the Landweber iteration with $k_*(\delta, y^\delta) = \mathcal{O}(\delta^{-2})$.*

Proof. Let us consider the sequence $\{x_k\}$ which corresponds to the Landweber iteration with the exact right-hand y. Following from the proof of Proposition 4.9, we have

$$\begin{aligned}
\|x^\dagger - x_j\|^2 - \|x^\dagger - x_{j+1}\|^2 &= \|y - Ax_j\|^2 + \langle y - Ax_j, (I - AA^*)(y - Ax_j)\rangle \\
&\geq \|y - Ax_j\|^2.
\end{aligned}$$

We then sum the inequalities from $j = 1$ through to k:

$$\begin{aligned} \|x^\dagger - x_1\|^2 - \|x^\dagger - x_{k+1}\|^2 &\geq \sum_{j=1}^{k} \|y - Ax_j\|^2 \\ &\geq k\|y - Ax_k\|^2, \end{aligned}$$

where the final inequality follows from the monotonicity of the residual norms. By induction and from equation (78), we have

$$y - Ax_k = (I - AA^*)^k (y - Ax_0),$$

and conclude

$$\|(I - AA^*)^k (y - Ax_0)\| = \|y - Ax_k\| \leq k^{-\frac{1}{2}} \|x^\dagger - x_1\|.$$

We now estimate the norm of the real residual $y^\delta - Ax_k^\delta$, we have

$$\begin{aligned} \|y^\delta - Ax_k^\delta\| &= \|(I - AA^*)^k (y^\delta - Ax_0)\| \\ &\leq \|(I - AA^*)^k (y^\delta - y)\| + \|(I - AA^*)^k (y - Ax_0)\| \\ &\leq \delta + k^{-\frac{1}{2}} \|x^\dagger - x_1\|. \end{aligned}$$

Consequently, the right-hand side is below $\eta\delta$ as soon as $k > (\eta - 1)^{-2} \|x^\dagger - x_1\|^2 \delta^{-2}$, and hence $k_*(\delta, y^\delta) \leq c\delta^{-2}$, where c depends on η only. This ends the proof. $\square$

REFERENCES

[1] Johann Baumeister. *Stable solution of inverse problems*. Springer, 1987.

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

[3] Mark Gockenbach. *Linear Inverse Problems and Tikhonov Regularization*. Number 32. The Mathematical Association of America, 2016.

[4] CW Groetsch. The theory of Tikhonov regularization for Fredholm equations. *104p, Boston Pitman Publication*, 1984.

[5] Charles W Groetsch and CW Groetsch. *Inverse problems in the mathematical sciences*, volume 52. Springer, 1993.

[6] CW Groetsch. Hofmann, b., Regularization for Applied Inverse and Ill-Posed Problems. Leipzig, BG Teubner 1986. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 68(12):654–654, 1988.

[7] Victor Isakov. *Inverse problems for partial differential equations*, volume 127. Springer, 2006.

[8] Andreas Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer Science & Business Media, 2011.

[9] Vladimir Alekseevich Morozov. *Methods for solving incorrectly posed problems*. Springer Science & Business Media, 2012.

[10] Anatoly Bakushinsky and A Goncharsky. *Ill-posed problems: theory and applications*, volume 301. Springer Science & Business Media, 2012.

[11] Michail M Lavrentiev. *Some improperly posed problems of mathematical physics*, volume 11. Springer Science & Business Media, 2013.

[12] Alfred Karl Louis. *Inverse und schlecht gestellte Probleme*. Springer-Verlag, 2013.

[13] Vladimir Alekseevich Morozov and Michael Stessin. *Regularization methods for ill-posed problems*. CRC press Boca Raton, FL:, 1993.

[14] Andrey N Tikhonov and Vasilii Iakkovlevich Arsenin. *Solutions of ill-posed problems*, volume 14. Winston, Washington, DC, 1977.

[15] GM Vainikko and A Yu Veretennikov. Iteration procedures in ill-posed problems, 1986.

[16] Joseph B. Keller. Inverse problems. *The American Mathematical Monthly*, 83(2):107–118, 1976.

[17] Jacques Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Courier Corporation, 2003.

[18] Jari P Kaipio and Erkki Somersalo. Computational and statistical methods for inverse problems. *Applied mathematical sciences*, 160, 2004.

[19] Martin Burger. Lecture notes on inverse problems, 2007. `https://www.uni-muenster.de/AMM/num/Vorlesungen/IP_WS07/skript.pdf` (visited: 2019-03-01).

[20] Ivan Tomba. *Iterative regularization methods for ill-posed problems*. PhD thesis, alma, 2013.

[21] Bastian von Harrach. Introduction to inverse problems, 2015. `https://www.math.uni-frankfurt.de/~harrach/talks/2015Seoul_Introduction_Inverse_Problems.pdf` (visited: 2019-05-01).

[22] Jennifer L Mueller and Samuli Siltanen. *Linear and nonlinear inverse problems with practical applications*, volume 10. Siam, 2012.

[23] Charles W Groetsch. Generalized inverses of linear operators: representation and approximation. 1977.

[24] Erwin Kreyszig. *Introductory functional analysis with applications*, volume 1. Wiley New York, 1978.

[25] A Bielicki. Une remarque sur la methode de Banach-Cacciopoli-Tikhonov. *Bull. Acad. Polon. Sci*, 4:261–268, 1956.

[26] LA Lusternik and VI Sobolev. Elements of Functional Analysis. 1974.

[27] Jinlu Li. The metric projection and its applications to solving variational inequalities in Banach spaces. *Fixed Point Theory*, 5(2):285–298, 2004.

[28] Michael Renardy and Robert C Rogers. *An introduction to partial differential equations*, volume 13. Springer Science & Business Media, 2006, pp. 273-274.

[29] Martin Benning. Inverse problems, updated on: June 6, 2016.

[30] Richard J Hanson. A numerical method for solving Fredholm integral equations of the first kind using singular values. *SIAM Journal on Numerical Analysis*, 8(3):616–622, 1971.

[31] Sergey I Kabanikhin. *Inverse and ill-posed problems: theory and applications*, volume 55. Walter De Gruyter, 2011.

[32] Rainer Kress, V Maz'ya, and V Kozlov. *Linear integral equations*, volume 82. Springer, 1989.

[33] Per Christian Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, volume 4. Siam, 2005.

[34] FR De Hoog. Review of Fredholm equations of the first kind. *The application and numerical solution of integral equations*, pages 119–134, 1980.

[35] Per Christian Hansen. Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank. *SIAM Journal on Scientific and Statistical Computing*, 11(3):503–518, 1990.

[36] Per Christian Hansen. Computation of the singular value expansion. *Computing*, 40(3):185–199, 1988.

[37] M Zuhair Nashed. *Generalized Inverses and Applications: Proceedings of an Advanced Seminar Sponsored by the Mathematics Research Center, the University of Wisconsin—Madison, October 8-10, 1973*. Number 32. Elsevier, 2014.

[38] CW Groetsch. Spectral methods for linear inverse problems with unbounded operators. *Journal of approximation theory*, 70(1):16–28, 1992.

[39] F Riesz and B Sz Nagy. Functional Analysis, Dover Publications, New York, 1990.

[40] Charles W Groetsch. The delayed emergence of regularization theory. *Bollettino di storia delle scienze matematiche*, 23:105–120, 2003.

[41] AB Bakushinskii. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *USSR Computational Mathematics and Mathematical Physics*, 24(4):181–182, 1984.

[42] Barbara Kaltenbacher, Andreas Neubauer, and Otmar Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6. Walter de Gruyter, 2008.

[43] Louis Landweber. An iteration formula for Fredholm integral equations of the first kind. *American journal of mathematics*, 73(3):615–624, 1951.

[44] VM Fridman. Method of successive approximations for a Fredholm integral equation of the 1st kind. *Uspekhi Matematicheskikh Nauk*, 11(1):233–234, 1956.

[45] Horst Bialy. Iterative behandlung linearer funktionalgleichungen. *Archive for Rational Mechanics and Analysis*, 4(1):166–176, 1959.

[46] Mario Bertero. Linear inverse and ill-posed problems. In *Advances in electronics and electron physics*, volume 75, pages 1–120. Elsevier, 1989.

[47] Sergios Theodoridis and Rama Chellappa. *Academic Press Library in Signal Processing: Image, Video Processing and Analysis, Hardware, Audio, Acoustic and Speech Processing*, volume 4. Academic Press, 2013.

[48] Frank Natterer. The mathematics of computerized tomography: A Wiley-Teubner publication. 1986.

[49] PC Hansen. Rank-Deficient and Discrete Ill-Posed Problems (philadelphia, pa: Siam). 1998.

[50] HW Engl, M Hanke, and A Neubauer. Regularization of inverse problems kluwer. *Dordrecht, The Netherlands*, 2000.

[51] Tommy Elfving and Touraj Nikazad. Stopping rules for landweber-type iteration. *Inverse Problems*, 23(4):1417, 2007.

[52] Martin Hanke, Andreas Neubauer, and Otmar Scherzer. A convergence analysis of the landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72(1):21–37, 1995.

[53] Vladimir Alekseevich Morozov. On the solution of functional equations by the method of regularization. In *Doklady Akademii Nauk*, volume 167, pages 510–512. Russian Academy of Sciences, 1966.

[54] Gennadii M Vainikko. Error estimates of the successive approximation method for ill-posed problems. *Avtomatika i Telemekhanika*, (3):84–92, 1980.

[55] Michel Defrise, Christine De Mol, and Pierre Célestin Sabatier. A note on stopping rules for iterative regularisation methods and filtered svd. *Inverse Problems: An Interdisciplinary Study*, pages 261–268, 1987.

[56] Ronny Ramlau. A modified landweber method for inverse problems. *Numerical Functional Analysis and Optimization*, 20(1-2):79–98, 1999.