Prediction bands for ill-posed problems
by Andrzej Wilhelm Jonca

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics
Montana State University

Abstract:
Prediction bands for regularized solutions to linear operator equations are constructed to assess the reliability of the solutions. These equations are ill-posed, i.e., small perturbations in the data may lead to large perturbations in the solution. To obtain an approximate solution, spectral filtering is used. The additive model is used and both the solution and noise in the data are assumed to be Gaussian stochastic processes. Numerical results are presented for the case of convolution integral operators.

PREDICTION BANDS FOR ILL-POSED PROBLEMS

by

Andrzej Wilhelm Jonca

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Mathematics

MONTANA STATE UNIVERSITY
Bozeman, Montana

July, 1988

APPROVAL

of a thesis submitted by

Andrzej Wilhelm Jonca

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

7/28/88
Date

Curtis R. Vogel
Chairperson,
Graduate Committee

Approved for the Major Department

7/28/88
Date

K. Taylor
Head, Major Department

Approved for the College of Graduate Studies

July 28, 1988
Date

Henry J. Parsons
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute copies of the dissertation in and from microfilm and the right to reproduce and distribute by abstract in any format."

Signature Andrzej W. Mynca

Date July 28, 1988

ACKNOWLEDGMENTS

I would like to thank Professor Curtis Vogel for the valuable discussions with me, constant care and encouragement during my work with him.

I also want to thank Professor Robert Boik and Professor John Lund for reading the text and offering constructive comments.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Prediction bands for regularized solutions to linear operator equations are constructed to assess the reliability of the solutions. These equations are ill-posed, i.e., small perturbations in the data may lead to large perturbations in the solution. To obtain an approximate solution, spectral filtering is used. The additive model is used and both the solution and noise in the data are assumed to be Gaussian stochastic processes. Numerical results are presented for the case of convolution integral operators.

CHAPTER 1

INTRODUCTION

This thesis deals with the linear ill-posed operator equation

$$Kx = z \tag{1.1}$$

where $K : \mathcal{X} \to \mathcal{Y}$ is an operator between two Hilbert spaces. The ill-posedness of the problem (see [3]) means that a small perturbation in the data z may result in large changes in the solution to (1.1). This discontinuous dependence of the solution on the data requires regularization in order to approximately solve the ill-posed equation.

The objective of this thesis is to provide a framework for analyzing the reliability of regularized solutions to problem (1.1). We will discuss a class of regularization methods, called spectral filtering methods (see [22]). To analyze these methods, consider the additive model for noisy data

$$z = Kx_{true} + \epsilon, \tag{1.2}$$

where x_{true} is the underlying true solution, which is defined on a set T, and ϵ is noise in the data. Assume x_{true} and ϵ are realizations of Gaussian stochastic processes with 0 means and known covariances.

We will quantify reliability by computing two types of "prediction bands" about the regularized solution:

(i) pointwise band: for each $t \in T$, true solution will be within this band with some prescribed probability, which we refer to as a confidence level.

(ii) uniform band (or "Scheffé band"): the probability that the entire solution will be within the band is given by the prescribed confidence level.

This thesis was partly motivated by work of O.N. Strand and E.R. Westwater in [19]. They assumed that both true solution and noise are Gaussian stochastic processes and obtained solutions to Fredholm integral equations of the first kind which are a special case of (1.1). Also G. Wahba in [12] obtained what she referred to as "Bayesian confidence intervals" based on the posterior covariance. The compact operator K in her paper was pointwise evaluations of functions in certain Hilbert spaces of "smooth" functions.

The organization of the thesis is as follows: Chapter 2 contains operator theory preliminaries. The ideas used in the sequel such as compact operators, Fourier transformation, the Fast Fourier Transform, generalized inverses, ill-posedness and singular value decomposition are reviewed. Chapter 3 summarizes the relevant probability concepts. Chapter 4 describes the error analysis, spectral filtering and construction of prediction bands. Then numerical results for special case of convolution integral operator are presented in Chapter 5.

CHAPTER 2

OPERATOR THEORY PRELIMINARIES

Hilbert Spaces

Throughout the thesis $\mathcal{X}$ will be a Hilbert space over the field of complex numbers $\mathbf{C}$ unless specified otherwise, with inner product denoted by

$$\langle x, y\rangle, \quad x, y \in \mathcal{X},$$

and induced norm

$$\|x\| := \sqrt{\langle x, x\rangle}, \quad x \in \mathcal{X}.$$

Here ":=" indicates a definition. Two specific examples of Hilbert spaces used in this thesis are:

(i) $L^2[a,b]$ is the set of all equivalence classes of functions $x : [a,b] \to \mathbf{C}$, which are square integrable, that is $\int_a^b |x^2(t)|\, dt$ is finite. The inner product is defined as

$$\langle x, y\rangle := \int_a^b x(t)\overline{y(t)}\, dt \quad x, y \in L^2[a,b]. \tag{2.1}$$

(ii) $H^p[a,b]$ is the set of all functions $x : [a,b] \to \mathbf{C}$ such that $\frac{d^p x}{dt^p} \in L^2[a,b]$. The standard inner product is defined as

$$\langle x, y\rangle := \sum_{k=0}^{p} \int_a^b \frac{d^k x}{dt^k}(t)\, \overline{\frac{d^k y}{dt^k}}(t)\, dt \quad x, y \in H^p[a,b]. \tag{2.2}$$

In particular, for $H^1[a,b]$ we have

$$\langle x, y\rangle = \int_a^b x(t)\overline{y(t)}\, dt + \int_a^b x'(t)\overline{y'(t)}\, dt. \tag{2.3}$$

Another inner product in $H^1[a,b]$ yielding a norm equivalent to the norm defined by (2.3) is

$$\langle x,y\rangle = x(a)\overline{y(a)} + \int_a^b x'(t)\overline{y'(t)}\,dt. \tag{2.4}$$

The subspace of $H^p[a,b]$ which consists of functions vanishing on the boundary will be denoted by $H_0^p[a,b]$. On the subspace $H_0^1[a,b]$, the inner product that will be used is

$$\langle x,y\rangle = \int_a^b x'(t)\overline{y'(t)}\,dt \quad x,y \in H_0^1[a,b]. \tag{2.5}$$

Definition 2.6: If $K : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces, is a continuous linear operator, then the adjoint operator of K will be denoted by K^*: $K^* : \mathcal{Y} \to \mathcal{X}$ and

$$\forall x \in \mathcal{X} \quad \text{and} \quad \forall y \in \mathcal{Y} \qquad \langle Kx,y\rangle = \langle x,K^*y\rangle.$$

Moreover, if $\mathcal{X} = \mathcal{Y}$ and $K^* = K$ then K is called a self-adjoint operator.

The Fourier Transformation

Definition 2.7: Let x be a function integrable on R: $x \in \mathcal{L}^1(R)$. The function

$$\hat{x}(\tau) := \int_R x(t)e^{-2\pi i t\tau}\,dt, \qquad \tau \in R, \quad i = \sqrt{-1},$$

is called the Fourier transform of x. Typically x is termed a function of time and $\hat{x}$ is termed a function of frequency. The mapping $\mathcal{F} : x \to \hat{x}$ is called the Fourier transformation:

$$\hat{x} = \mathcal{F}x. \tag{2.8}$$

Some results to be applied later are now reviewed.

If $x \in \mathcal{L}^1(R)$ is infinitely differentiable and has a compact support, then integrating by parts p times, we arrive at the formula

$$(\mathcal{F}\frac{d^p x}{dt^p})(\tau) = (2\pi i)^p \tau^p (\mathcal{F}x)(\tau). \tag{2.9}$$

Clearly the Fourier transformation is a linear operator. For several reasons (see [5]) it is inconvenient to have the space $\mathcal{L}^1(R)$ as its domain.

Definition 2.10: The space J is the set of all functions $x \in C^\infty(R)$ such that

$$\sup_{t \in R} |t^r \frac{d^p x}{dt^p}(t)| < \infty \qquad \text{for all } t \text{ and all nonnegative integers } p \text{ and } r.$$

The following statements hold:

Theorem 2.11:

(i) $J \subset \mathcal{L}^1(R)$,

(ii) the Fourier transformation $\mathcal{F} : J \to J$ is continuous,

(iii) J is dense in $L^2(R)$ (if the function is identified with the equivalence class it represents),

(iv) $\mathcal{F}$ is an isometry of $L^2(R)$ onto $L^2(R)$,

(v) if $x, y \in J$, then

$$\widehat{x * y} = \hat{x}\hat{y} \quad \text{and} \quad \widehat{xy} = \hat{x} * \hat{y}.$$

Here $x * y$ is the convolution of x and y, that is

$$(x * y)(s) := \int_{-\infty}^{\infty} x(s-t)y(t)\,dt.$$

Proof: See [5].

If x is a periodic function, then $\hat{x}$ may be properly defined only if one introduces the theory of distributions. An elementary discussion of it can be found in [10], and more detailed treatment in [5] or [11]. The distribution theory is also a natural tool to develop the discrete Fourier transformation described briefly in the next section.

Discrete Fourier Transformation

To determine the Fourier transformation of x computationally one needs to work with finite sequences representing x and its Fourier transform $\hat{x}$. The discrete Fourier transform pair approximates the original Fourier transform pair. The results of a theoretical development, which can be found in [10], are as follows:

Let $x(jh)$, $j = 0, 1, \ldots, n-1$ be a discrete version of x (x must be thought of as a periodic function here and the points $0, h, \ldots, (n-1)h$ are within one period of x). Then

$$\hat{x}(k/nh) = \sum_{j=0}^{n-1} x(jh) e^{-\frac{2\pi ijk}{n}}, \qquad x(jh) = \frac{1}{n} \sum_{k=0}^{n-1} \hat{x}(k/nh)\, e^{\frac{2\pi ijk}{n}},$$

where $k = 0, 1, \ldots, n-1$. This may also be expressed as

$$\hat{x}_d = \mathcal{F}_d x_d, \quad x_d = \mathcal{F}_d^{-1} \hat{x}_d \tag{2.12}$$

where $x_d, \hat{x}_d$ are discrete versions of x and $\hat{x}$, respectively. $\mathcal{F}_d$ and $\mathcal{F}_d^{-1}$ are matrices such that

$$[\mathcal{F}_d]_{jk} = \exp\left[\frac{2\pi ijk}{n}\right], \quad 0 \le j, k \le n-1,$$

$$\left[\mathcal{F}_d^{-1}\right]_{jk} = \frac{1}{n} \exp\left[-\frac{2\pi ijk}{n}\right], \quad 0 \le j, k \le n-1.$$

The Fast Fourier Transform (FFT) is an algorithm that rapidly computes the discrete Fourier transform. The FFT is used to obtain the numerical results

presented in this thesis. In the sequel the discrete Fourier transform and the inverse discrete Fourier transform will be referred to as FFT and IFFT, respectively.

Compact Operators

Definition 2.13: A set $M \subset \mathcal{X}$ is called compact if every sequence x_n of elements from M has a subsequence x_{n_k} converging to an element $x \in M$. A set M is called relatively compact if its closure $\overline{M}$ is compact.

Definition 2.14: A linear operator $K : \mathcal{X} \to \mathcal{Y}$ is called compact if for every bounded subset $M \subset \mathcal{X}$, the image $K(M)$ is relatively compact.

Example 2.15: If k is any square integrable function, that is

$$\int_a^b \int_a^b |k(s,t)|^2 \, ds \, dt,$$

is finite, then a typical example of a compact operator is

$$K : L^2[a,b] \to L^2[a,b]$$

where for $x \in L^2[a,b]$

$$(Kx)(s) = \int_a^b k(s,t)x(t) \, dt.$$

For a proof see [1].

Example 2.16: Let $K : H^p[a,b] \to L^2[a,b]$ and for $x \in H^p[a,b]$

$$(Kx)(s) = \int_a^b k(s,t)x(t) \, dt,$$

where k is a square integrable function. Then K is compact. To prove it, notice that K can be expressed as the composition $K = \tilde{K}J$, where $J : H_0^p[0,1] \to L^2[0,1]$

is an embedding and $\tilde{K} : L^2[0,1] \to L^2[0,1]$ is an operator from the previous example. Since J is continuous (and even compact — see [13]), K is compact.

In both Examples (2.15) and (2.16) an important case is the situation where

$$k(s,t) := k(s-t). \tag{2.17}$$

This kind of a kernel is called a convolution kernel. Its properties and applications will be discussed in detail later.

Theorem 2.18: (Spectral Theorem for Compact Self-Adjoint Operators).

Let $K : \mathcal{X} \to \mathcal{X}$ be a self-adjoint compact operator. Then K has the representation

$$Kx = \sum_{j \in J} \lambda_j \langle x, v_j \rangle v_j, \tag{2.19}$$

where the λ_j are eigenvalues of K (each repeated in the sum according to its multiplicity), the v_j are corresponding orthonormal eigenvectors, and J is an index set for the eigenvalues. J is countable. If J is infinite, 0 is the only limit point of the spectrum of K.

Proof: See [2].

Throughout the thesis the spectrum of a linear operator K will be denoted by $\sigma(K)$.

Singular Value Decomposition

Let $K : \mathcal{X} \to \mathcal{Y}$ be a compact operator. Then K^*K is both compact and self-adjoint. Denote by v_j the orthonormal eigenvectors of K^*K, that is

$$K^*Kv_j = \lambda_j v_j, \qquad \langle v_i, v_j \rangle = \delta_{ij}. \tag{2.20}$$

It is immediate that all eigenvalues λ_j are nonnegative:

$$\lambda_j = \lambda_j \langle v_j, v_j \rangle = \langle K^* K v_j, v_j \rangle = \langle K v_j, K v_j \rangle \geq 0.$$

Hence one can introduce the singular values σ_j of the operator K by:

$$\sigma_j := \sqrt{\lambda_j}, \qquad \text{for } \lambda_j > 0. \tag{2.21}$$

It will be assumed throughout the thesis that the singular values are ordered so that $\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots$.

Next define

$$u_j := \frac{1}{\sigma_j} K v_j. \tag{2.22}$$

A number of easy formulas follows:

$$K K^* u_j = \lambda_j u_j, \qquad \langle u_j, u_k \rangle = \delta_{jk}, \qquad \frac{1}{\sigma_j} K^* u_j = v_j.$$

The family $\{v_j, u_j, \sigma_j\}_{j \in J}$ is called a singular system for K. Notice that because K is compact, the index set is countable.

Definition 2.23: For $K : \mathcal{X} \to \mathcal{Y}$ the range of K, or the image of $\mathcal{X}$ under K, will be denoted by $\mathcal{R}(K)$.

$$\mathcal{R}(K) := \{y \in \mathcal{Y};\ \exists x \in \mathcal{X} \quad Kx = y\}.$$

The null space (kernel) of K is the inverse image of $0 \in \mathcal{Y}$ under K:

$$\text{Null}\, K := \{x \in \mathcal{X};\ \ Kx = 0\}.$$

Using the Spectral Theorem 2.18 one shows the following facts:

(i) $\overline{\text{Span}\,(u_j)_{j\in J}} = \overline{\mathcal{R}(K)}$,

(ii) $\overline{\text{Span}\,(v_j)_{j\in J}} = \overline{\mathcal{R}(K^*)}$.

From the standard identities:

$$\overline{\mathcal{R}(K^*)} = (\text{Null}\,K)^\perp, \qquad \overline{\mathcal{R}(K)} = (\text{Null}\,K^*)^\perp,$$

one obtains that $\text{Null}\,(K^*K) = \text{Null}\,K$, so finally

$$\forall x \in \mathcal{X} \quad x = x_0 + \sum_{j\in J} c_j v_j, \qquad \text{where } x_0 \in \text{Null}\,K.$$

Now we can obtain the singular value decomposition for K:

$$Kx = K(x_0 + \sum_{j\in J} c_j v_j) = \sum_{j\in J} c_j K v_j, \qquad \text{so}$$

$$Kx = \sum_{j\in J} c_j \sigma_j u_j = \sum_{j\in J} \sigma_j \langle x, v_j\rangle u_j. \tag{2.24}$$

If K is an $m \times n$ matrix then its singular value decomposition (SVD) is

$$K = UDV^*. \tag{2.25}$$

The columns of V are the singular vectors v_j. The columns of U are the singular vectors u_j. Both matrices V and U are Hermitian. The singular values σ_j lie on the main diagonal of the diagonal matrix D.

In Chapter 4 and 5 we will need the results of the following

Example 2.26: Singular system for an integral operator with a convolution kernel.

Consider first $K : L^2[0,1] \to L^2[0,1]$. It is well known that (see [1]) if $(Kx)(s) = \int_0^1 k(s,t)x(t)\,dt$, $(K^*y)(t) = \int_0^1 k^*(t,s)y(s)\,ds$, then $k^*(t,s) = \overline{k(s,t)}$.

For the convolution kernel $k(s,t) = k(s-t)$, assuming k real and periodic, and taking $x(t) = e^{2\pi int}$ we get (denote $u := s - \tau$):

$$\begin{aligned}(K^*Kx)(t) &= \int_0^1 k(s-t)[\int_0^1 k(s-\tau)e^{2\pi in\tau}\,d\tau]\,ds \\ &= \int_0^1 k(s-t)[\int_{s-1}^{s} k(u)e^{-2\pi inu}e^{2\pi ins}\,du]\,ds \\ &= \int_0^1 k(s-t)e^{2\pi ins}[\int_0^1 k(u)e^{-2\pi inu}\,du]\,ds \\ &= k_n\int_0^1 k(s-t)e^{2\pi ins}\,ds\end{aligned}$$

where $k_n := \int_0^1 k(u)e^{-2\pi inu}\,du$ is the Fourier coefficient of k. An identical change of variables shows that

$$\int_0^1 k(s-t)e^{2\pi ins}\,ds = \overline{k_n}e^{2\pi int}$$

hence

$$(K^*Kx)(t) = |k_n|^2 x(t). \tag{2.27}$$

Now consider K as an operator from $H_0^1[0,1]$ into $L^2[0,1]$ rather than that on $L^2[0,1]$, with other hypotheses unchanged. Applying the Definition 2.6 of an adjoint operator and also using (2.5) we easily obtain

$$\forall x \in H_0^1[0,1] \quad \int_0^1 x'(t)\frac{\partial k^*}{\partial t}(t,s)\,dt = \int_0^1 k(s,t)x(t)\,dt.$$

Integrating by parts and noticing that k^* is periodic, for $x(t) = e^{2\pi int}$, we arrive at

$$\int_0^1 x''(t)k^*(t,s)\,dt = -\int_0^1 k(s,t)x(t)\,dt, \tag{2.28}$$

and hence

$$(K^*Kx)(t) = k_n\int_0^1 k^*(s,t)e^{2\pi ins}\,ds.$$

Finally, using (2.28), the following result appears:

$$(K^*Kx)(t) = -k_n \int_0^1 k(s-t)\frac{1}{(2\pi in)^2}e^{2\pi ins}\,ds = \frac{1}{(2\pi n)^2}|k_n|^2 x(t).$$

Analogous results hold for $p = 2, 3, \ldots$. Summarizing the Example 2.26 we can say that an integral operator $K : H_0^p[0,1] \to L^2[0,1]$ has a singular system

$$v_n(t) = e^{2\pi int}, \quad u_n(t) = e^{-2\pi int}, \quad \sigma_n = \frac{|k_n|}{(2\pi n)^p}, \tag{2.29}$$

(the results for u_n are derived in the same manner as for v_n).

There is a relationship between the matrices $\mathcal{F}_d, \mathcal{F}_d^{-1}$ of the discrete Fourier transformation (see (2.12)) and the singular value decomposition of the matrix K representing the discretized version of an integral operator with convolution kernel. According to (2.25)

$$v_k(t_j) = \frac{1}{\sqrt{n}}e^{\frac{2\pi ijk}{n}}, \qquad u_k(t_j) = \frac{1}{\sqrt{n}}e^{-\frac{2\pi ijk}{n}}.$$

where $\frac{1}{\sqrt{n}}$ is a normalizing factor so that V and U that have the v_k and u_k as their columns, respectively, are Hermitian matrices. We obtain formulas needed in Chapter 4 and 5:

$$U = \frac{1}{\sqrt{n}}\mathcal{F}_d, \qquad V = \sqrt{n}\mathcal{F}_d^{-1}. \tag{2.30}$$

Moore-Penrose Generalized Inverse

A classical solution to an operator equation (1.1) exists if and only if $y \in \mathcal{R}(K)$. We introduce a concept of a generalized solution — a least squares solution.

Definition 2.31: The set of least squares solutions to $Kx = y$ is defined by

$$S_y := \{u \in \mathcal{X}; \quad \forall x \in \mathcal{X} \quad \|Ku - y\| \le \|Kx - y\|\}.$$

Note that S_y may be empty. If S_y contains an element x_0, then

$$S_y = \{x_0\} + \text{Null}\, K.$$

S_y is closed and convex. It can consist of a single element only if $\text{Null}\, K = \{0\}$. If $S_y \neq \emptyset$, we define the least squares minimum norm solution $x \in S_y$ by

$$\forall u \in S_y \quad \|x\| \leq \|u\|.$$

Definition 2.32: The Moore-Penrose generalized inverse operator

$$K^\dagger : \mathcal{D}(K^\dagger) \subset \mathcal{Y} \to \mathcal{X}$$

is given by

$$\mathcal{D}(K^\dagger) = \{y \in \mathcal{Y}; \quad S_y \neq \emptyset\}$$

and $K^\dagger y$ is the least squares minimum norm solution.

Theorem 2.33:

(i) $\mathcal{D}(K^\dagger) = \mathcal{R}(K) \oplus (\mathcal{R}(K))^\perp$,

(ii) $\mathcal{D}(K^\dagger)$ is dense in $\mathcal{Y}$,

(iii) $\mathcal{R}(K^\dagger) \subset (\text{Null}\, K)^\perp$.

Proof: See [2].

The following representation for the generalized inverse is frequently used in this thesis:

Theorem 2.34: For any $y \in \mathcal{D}(K^\dagger)$

$$K^\dagger y = \sum_{j \in J} \frac{\langle y, u_j \rangle}{\sigma_j} v_j.$$

Proof: See [3].

The next two theorems show that except in simple cases $K^\dagger$ is not continuous.

Theorem 2.35: Let K be a compact operator. Then

$$\left(K^\dagger \text{ is continuous}\right) \iff \left(\mathcal{R}(K) \text{ is closed in Y}\right).$$

Proof: See [2].

Theorem 2.36: Let K be a compact operator. Then

$$\left(K^\dagger \text{ is continuous}\right) \iff \left(\dim \mathcal{R}(K) < \infty\right).$$

Proof: $\Rightarrow$ Notice that $KK^\dagger = I\,\big|_{\mathcal{R}(K)}$. Because $K^\dagger$ is continuous and K is compact, $KK^\dagger$ is compact. By the Riesz Lemma (see [4]) an identity operator is compact if and only if its domain is finite dimensional.

$\Leftarrow$ Any finite dimensional subspace is closed (see [4]), hence $\mathcal{R}(K)$ is closed. By Theorem 2.35 $K^\dagger$ is continuous.

Ill-Posedness

Definition 2.37: Let $K : \mathcal{X} \to \mathcal{Y}$. The problem $Kx = y$ is well-posed provided that the following three conditions hold:

(i) $\forall y \in \mathcal{Y}$, there a solution $x \in \mathcal{X}$,

(ii) the solution x is unique,

(iii) the solution x depends continuously on data y.

The problem is called ill-posed if it is not well-posed.

Theorem (2.36) shows that except in trivial cases of finite rank operators ($\mathcal{R}(K)$ finite dimensional) the equation $Kx = y$ with K compact is ill-posed even

if a solution to the problem is taken to be the least squares minimum norm solution. The discontinuous dependence on the data is an inherent feature of the operator K. Obviously ill-posedness depends on the choice of the Hilbert spaces X and Y. Physical considerations often dictate the choice of the spaces for which many practical problems become ill-posed.

Sinc Quadrature Formula

In order to introduce the quadrature theorem that is used in the thesis, the following concepts need to mentioned:

Let f be a function analytic in a simply connected domain $\mathcal{D} \subset \mathbf{C}$. f must satisfy two technical conditions — a detailed description can be found in [14] or [15]. Let ϕ be a conformal (that is, for every $z \in \mathcal{D}$, there exists $\phi'(z) \neq 0$) bijection of $\mathcal{D}$ onto S_d, an infinite strip of width $2d$ about the real axis.

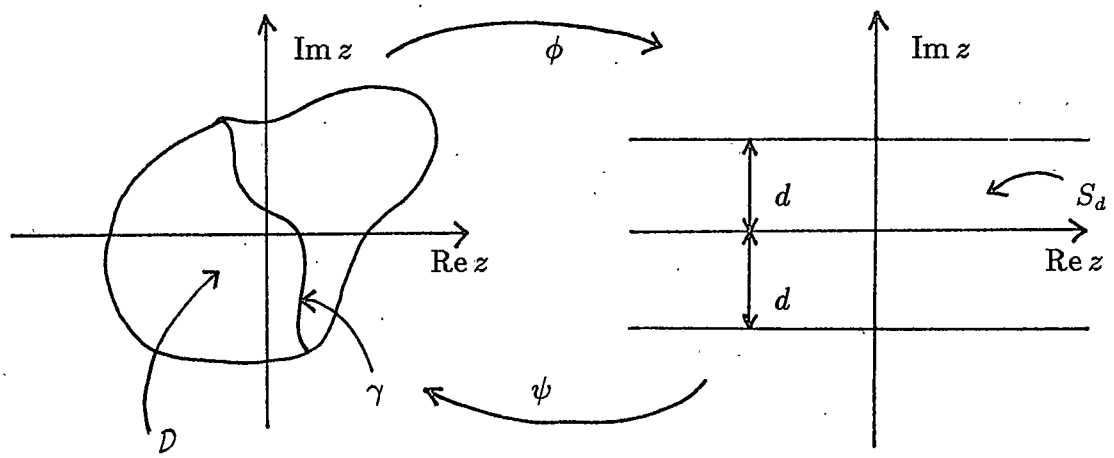

Figure 1: Bijection ϕ of $\mathcal{D}$ onto S_d.

Let $\psi := \phi^{-1}$, $\gamma := \psi(\mathbf{R})$, $\gamma_L := \psi(-\infty, 0)$, $\gamma_R := \psi(0, \infty)$. Then, if f

satisfies the following growth condition:

$$\left|\frac{f(z)}{\phi'(z)}\right| \leq const \begin{cases} e^{-\alpha|\phi(z)|}, & z \in \gamma_L; \\ e^{-\beta|\phi(z)|}, & z \in \gamma_R. \end{cases} \tag{2.38}$$

the following inequality holds:

$$\left|\int_\gamma f(z)\,dz - h\sum_{k=-M}^{N} \frac{f(z_k)}{\phi'(z_k)}\right| \leq const\ e^{-(2\pi d\beta N)^{1/2}} \tag{2.39}$$

where $h = \left(\frac{2\pi d}{\beta N}\right)^{1/2}$, $M = \left[\frac{\beta}{\alpha}N + 1\right]$, $z_k = \psi(kh)$.

An integral that is computed in the thesis has the form

$$I = \int_0^\infty \frac{\sin\{\frac{1}{2}[\sum_{j\in J}\arctan(c_j t) - xt]\}}{t\prod_{j\in J}(1+c_j t^2)^{\frac{1}{4}}}\,dt$$

It poses difficulties because of the infinite range of integration and an oscillatory integrand. If ϕ is chosen to be:

$$\phi(z) := \log(\sinh z),$$

the following quadrature is obtained:

$$\int_0^\infty f(x)\,dx = h\sum_{k=-M}^{N} \frac{1}{\sqrt{1+e^{-2kh}}} f(\log(e^{kh} + \sqrt{e^{2kh}+1})).$$

The condition (2.38) becomes:

$$|f(x)| \leq const \begin{cases} x^{\alpha-1}, & x \in (0, \log(1+\sqrt{2})); \\ e^{-\beta x}, & x \in (\log(1+\sqrt{2}), \infty). \end{cases}$$

Clearly the integrand involved does not decay exponentially; however this affects only the rate of convergence and the choice of h, M and N. There are three contributions to the quadrature error: approximating the integrand, truncating the infinite sum $\sum_{k=-\infty}^{\infty} \frac{f(z_k)}{\phi'(z_k)}$ below, and truncating it above. Balancing the different errors so that asymptotically they are identical leads to

$$h = \frac{\pi}{\sqrt{M}} \quad \text{and} \quad N = \frac{\sqrt{M}}{\pi} e^{\frac{2}{n}\pi M}.$$

With these selections the rate of convergence in (2.39) is maintained. The value of d is taken to be $d = \frac{\pi}{2}$. This ensures that ϕ is conformal.

CHAPTER 3

STATISTICAL PRELIMINARIES

Random Variables and their Distributions

Let (S, E, P) denote a probability space, i.e., let S denote the sample space, let E denote the family of events, and let P denote a probability measure defined on S (see [20]).

Definition 3.1: A real valued function X defined on the sample space S is called a random variable if for every Borel set $B \subset R$, the set $\{s \in S;\ X(s) \in B\} \in E$, that is, it is an event in E.

Definition 3.2: The cumulative distribution function (CDF) of a random variable X is a function $F_X : R \to R$ defined by

$$F_X(x) := P(X \leq x).$$

In case of a continuous random variable its CDF can be represented as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt,$$

where f_X is called the probability density function (pdf).

Definition 3.3: The joint cumulative distribution function of the n random variables $X_1, X_2, \cdots, X_n$ is defined by

$$F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n).$$

In the continuous case

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\cdots,X_n}(t_1,\cdots,t_n)\, dt_n \cdots dt_1.$$

Definition 3.4: Random variables $X_1, X_2, \cdots, X_n$ are said to be independent if

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_n).$$

Definition 3.5: The expected value, or mean, $\mathcal{E}(X)$ of a continuous random variable X is defined by

$$\mathcal{E}(X) := \int_R x f_X(x)\, dx, \qquad \text{whenever the integral converges.}$$

The variance of X is

$$\mathrm{Var}(X) := \mathcal{E}(X - \mathcal{E}(X))^2 = \mathcal{E}(X^2) - (\mathcal{E}(X))^2.$$

The covariance of two random variables X and Y is defined by

$$\mathrm{Cov}(X,Y) := \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y).$$

Definition 3.6: The characteristic function φ_X of a random variable X is defined by

$$\varphi_X(t) := \mathcal{E}(e^{itx}) = \int_R e^{itx} f_X(x)\, dx.$$

Note that ϕ_X is the inverse Fourier transform of its probability density f_X. There is one-to-one correspondence between cumulative distribution functions and characteristic functions. For independent random variables we have

$$\varphi_{X_1,\cdots,X_n}(t_1,\cdots,t_n) = \varphi_{X_1}(t_1)\cdots\varphi_{X_n}(t_n).$$

Now let A denote an $m \times n$ matrix whose elements are random variables A_{ij}. Define

$$\mathcal{E}(A) := \begin{pmatrix} \mathcal{E}(A_{11}) & \dots & \mathcal{E}(A_{1n}) \\ \vdots & & \vdots \\ \mathcal{E}(A_{m1}) & \dots & \mathcal{E}(A_{mn}) \end{pmatrix}.$$

In particular consider the random vector $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ and let $\mu := \mathcal{E}(\mathbf{X}) = \begin{pmatrix} \mathcal{E}(X_1) \\ \vdots \\ \mathcal{E}(X_n) \end{pmatrix}$.

Definition 3.7: The covariance matrix M is

$$M := \mathcal{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^*].$$

Notice that

(i) $m_{ij} := [M]_{ij}$ is the covariance of X_i and X_j.

(ii) $M = M^*$ and M is positive semi-definite, that is, $\forall c \quad c^* M c \geq 0$.

(iii) $m_{jj} = \text{Var}(X_j)$.

(iv) If $X_1, X_2, \cdots, X_n$ are independent, the covariance matrix is diagonal.

Two important distributions of random variables are used in the thesis: the normal (or Gaussian) distribution and the chi-square distribution. Recall that a random variable X follows the normal distribution with mean μ and variance σ^2 if it has the following pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.8}$$

We write $X \sim \mathcal{N}(\mu, \sigma^2)$.

The following properties of a normally distributed random variable are used:

(i) if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z := \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1), \tag{3.9}$$

(ii) if $\mathbf{X}$ is a random vector normally distributed as $\mathbf{X} \sim \mathcal{N}(\mu, B)$ and A is a linear transformation, then (see [21])

$$A\mathbf{X} \sim \mathcal{N}(A\mu, ABA^*). \tag{3.10}$$

In particular, taking

$$A = (c_1 \quad c_2 \quad \dots \quad c_n),$$

we arrive at the conclusion that if $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$; $j = 1, 2, \dots, n$ denote independent normal variables, then

$$Y = \sum_{j=1}^{n} c_j X_j \sim \mathcal{N}(\sum_{j=1}^{n} c_j \mu_j, \sum_{j=1}^{n} c_j^2 \sigma_j^2).$$

Now recall that a random variable X has a gamma distribution with parameters $\kappa > 0$ and $\theta > 0$ if it has pdf of the form

$$f(x) = \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-\frac{x}{\theta}}, \qquad x > 0.$$

A special case of the gamma distribution with $\theta = 2$ and $\kappa = \frac{\nu}{2}$ is called a chi-square distribution with ν degrees of freedom. We write $X \sim \chi^2(\nu)$. We have the following remark

Remark 3.11:

(i) $\mathcal{E}(X) = \nu$,

(ii) $\mathrm{Var}(X) = 2\nu$,

(iii) if $Z \sim \mathcal{N}(0,1)$ then $Z^2 \sim \chi^2(1)$; more generally, if $\mathbf{X}$ is a random vector normally distributed with mean vector 0 and covariance matrix B, then the

quadratic form $Y = \mathbf{X}^* A\mathbf{X}$ is distributed as a linear combination of independent chi-square random variables, each with one degree of freedom:

$$Y \sim \sum_{j \in J} c_j \chi^2(1). \tag{3.12}$$

The coefficients c_j of the linear combination are eigenvalues of the matrix AB (see [6]).

The CDF of the random variable Y in (3.12) can be obtained using an inversion formula for the characteristic function φ_Y of the variable Y

$$\varphi_Y(t) = \prod_{j \in J} (1 - 2ic_j t)^{-\frac{1}{2}},$$

$$F_Y(y) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty t^{-1} Im\{e^{-ity} \varphi(t)\}\, dt.$$

Then one can show (see [6]) that

$$F_Y(y) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin\{\frac{1}{2}[\sum_{j \in J} \arctan(c_j t) - yt]\}}{t \prod_{j \in J} (1 + c_j t^2)^{\frac{1}{4}}}\, dt. \tag{3.13}$$

Stochastic Processes

Definition 3.14: A stochastic process is a family of random variables $(X(t))_{t \in T}$ defined on a common probability space.

A stochastic process $(X(t))_{t \in T}$ can be viewed as a function of two arguments $(X(t,s))_{t \in T, s \in S}$. For a fixed value of t, $X(t, \cdot)$ is a function on the sample space S, that is, $X(t, \cdot)$ is a random variable. On the other hand, for fixed s, $X(\cdot, s)$ is a function of t that represents a possible observation of the stochastic process. We say that $X(\cdot, s)$ is a realization of the process, or a sample function of the process.

The role played for a single random variable by its mean and variance is played for a stochastic process by respectively its mean value function $\mu(t) := \mathcal{E}(X(t))$ and its covariance function $K(t_1, t_2) := \mathrm{Cov}[X(t_1), X(t_2)]$, $t_1, t_2 \in T$.

Definition 3.15: A stochastic process $(X(t))_{t\in T}$ is called Gaussian if every linear combination of the random variables $X(t), t \in T$, is normally distributed. When the random variables $X(t_1), X(t_2), \cdots, X(t_n)$ have a joint normal distribution (which exists if and only if the covariance matrix of $X_1, X_2, \cdots, X_n$ is nonsingular), then the stochastic process $(X(t))_{t\in T}$ is Gaussian if and only if for all subsets $\{t_1, t_2, \cdots, t_n\} \subset T$, the random variables $X(t_1), X(t_2), \cdots, X(t_n)$ are jointly normally distributed.

More about stochastic processes can be found in [17].

Statistical Model

Consider the model (1.2) for noisy data. It is assumed that the true solution x_{true} is a realization of a Gaussian stochastic process $X(t), t \in T$, and the error ϵ is a realization of an independent Gaussian stochastic process: $\epsilon = (\epsilon_i)_{i=1}^n$. The stochastic form of (1.2) is

$$Z = KX + \epsilon. \tag{3.16}$$

We assume that $\mathcal{E}(X(t)) = 0$ and $\mathcal{E}(\epsilon) = 0$. If $\mathcal{E}(X(t))$ is not zero, (3.16) can always be "rescaled" in the following way: let $\bar{X}(t) := \mathcal{E}(X(t))$, $\bar{Z} := K\bar{X}$, $\tilde{X} := X - \bar{X}$, $\tilde{Z} := Z - \bar{Z}$. Then $K\tilde{X} = KX - K\bar{X} = Z - \bar{Z} = \tilde{Z}$, and we have a problem $K\tilde{X} = \tilde{Z}$ where $\mathcal{E}(\tilde{X}(t)) = 0$.

Similarly, if $\mathcal{E}(\epsilon(t)) \neq 0$, an analogous procedure can be done.

The stochastic model (3.16) can be constructed in the following way: We assume independent normally distributed random vectors

$$X \sim \mathcal{N}(0, C_x), \qquad \epsilon \sim \mathcal{N}(0, C_\epsilon).$$

These can be obtained by taking:

$$X := B_x \xi, \qquad \epsilon := B_\epsilon \eta, \tag{3.17}$$

where $C_x := B_x B_x^*$, $C_\epsilon := B_\epsilon B_\epsilon^*$, and

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \right). \tag{3.18}$$

A prediction problem (see [18]) can now be formulated: given a realization of the data z, predict what realization x gave rise to it. One can show ([18], [21]) that, by minimizing the squared error loss function

$$L(X, \hat{X}) = \mathcal{E}[\|X - \hat{X}\|^2],$$

one obtains that the best unbiased predictor of x is

$$\hat{x} = \mathcal{E}(X|Z = z) = C_x K^* [K C_x K^* + C_\epsilon]^{-1} z. \tag{3.19}$$

CHAPTER 4

REGULARIZATION AND THE CONSTRUCTION OF PREDICTION BANDS

Spectral Filtering

Consider the equation

$$Kx = z \tag{4.1}$$

with $K : \mathcal{X} \to \mathcal{Y}$ a compact operator. In practice we have noise contaminated data.

$$z := Kx_{true} + \epsilon \tag{4.2}$$

where ϵ represents noise in the data and x_{true} represents the underlying "true" solution. We know from Theorem 2.34 that if $z \in \mathcal{D}(K^\dagger)$, then a least squares minimum norm solution to (4.1) exists and can be expressed as

$$K^\dagger z = \sum_{j \in J} \frac{\langle z, u_j \rangle}{\sigma_j} v_j. \tag{4.3}$$

Let $P : \mathcal{X} \to (\text{Null}\, K)^\perp \subset \mathcal{X}$, denote the orthogonal projection of $\mathcal{X}$ onto $(\text{Null}\, K)^\perp$. The projection of the true solution onto the orthogonal complement of the null space of K is

$$Px_{true} = K^\dagger K x_{true} = \sum_{j \in J} \langle x_{true}, v_j \rangle v_j.$$

Assuming that $\epsilon \in \mathcal{D}(K^\dagger)$ the least squares minimum norm solution of $Kx = z$ is given by $K^\dagger z$. The difference between the two solutions becomes:

$$K^\dagger z - Px_{true} = K^\dagger(Kx_{true} + \epsilon) - K^\dagger K x_{true} = K^\dagger \epsilon = \sum_{j \in J} \frac{\langle \epsilon, u_j \rangle}{\sigma_j} v_j.$$

When K has infinite dimensional range, then the last expression shows conspicuously the ill-posedness of the problem: even if $\|\epsilon\|$ is small, that is, one set of the data differs little from the other one, because $\sigma_j \to 0$ as $j \to \infty$, $\|K^\dagger z - Px_{true}\|$ may be arbitrarily large.

To overcome this difficulty let us consider regularized solutions to (4.1) of the form

$$x_\alpha := \sum_{j \in J} w_j(\alpha) \frac{\langle z, u_j \rangle}{\sigma_j} v_j. \tag{4.4}$$

The $w_j(\alpha)$ are called weights and the sequence $(w_j(\alpha))_{j \in J}$ is called a spectral filter. The nonnegative real number α is referred to as a regularization parameter. Every spectral filter function $w(\sigma, \alpha)$, where $w(\sigma_j, \alpha) = w_j(\alpha)$, should possess the following characteristic:

$$w(\sigma, \alpha) \approx \begin{cases} 1, & \text{when } \sigma \text{ is "large"}; \\ 0, & \text{when } \sigma \text{ is "small"}. \end{cases}$$

Example 4.5: The truncated SVD filter:

$$w(\sigma, \alpha) = \begin{cases} 1, & \text{for } \sigma \geq \alpha; \\ 0, & \text{for } \sigma < \alpha. \end{cases}$$

where α is called a truncation level. In this way the amount of filtering increases with α increasing. This kind of filter is described in [7].

Example 4.6: The Tikhonov filter:

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

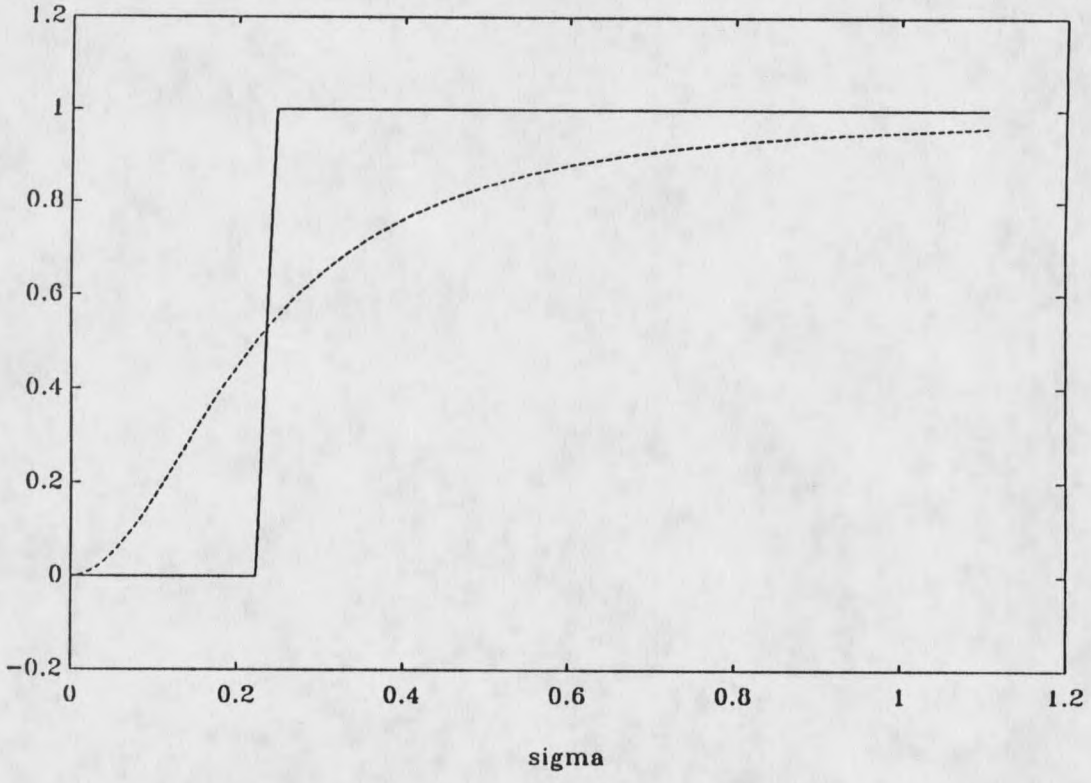

Figure 2. Spectral filtering functions $w(\sigma, \alpha)$ for Tikhonov regularization (dashed line) and TSVD (solid line) plotted as functions of σ; $\alpha = 0.05$ is fixed.

Again the bigger the α the more spectral filtering is applied, whereas $\alpha = 0$ corresponds to no spectral filtering at all.

The Tikhonov filter has the following variational characterization. If the Tikhonov functional is defined as

$$f_\alpha(x) := \frac{1}{2}\{\|Kx - z\|^2 + \alpha\|x\|^2\}, \quad x \in \mathcal{X}$$

where $\mathcal{X}$ is a Hilbert space, then a necessary condition for f_α to have a minimum

at x_α is

$$\forall h \in \mathcal{X} \quad f'_\alpha(x_\alpha)h = 0.$$

Since

$$f'_\alpha(x_\alpha)h = \langle Kx_\alpha - z, Kh\rangle + \alpha\langle x_\alpha, h\rangle = \langle K^*(Kx_\alpha - z) + \alpha I, h\rangle$$

it follows that

$$K^*(Kx_\alpha - z) + \alpha I = 0 \quad \text{or} \quad x_\alpha = (K^*K + \alpha I)^{-1}K^*z.$$

Notice that since all eigenvalues λ_j of K^*K are nonnegative (see 2.20) and $\alpha > 0$, the operator $K^*K + \alpha I$ is certainly invertible. Moreover, from Spectral Mapping Theorem (see [4]), (2.19) and (2.22) we have

$$\begin{aligned} x_\alpha &= \sum_{j\in J} \frac{1}{\lambda_j + \alpha}\langle K^*z, v_j\rangle v_j = \sum_{j\in J} \frac{\sigma_j}{\sigma_j^2 + \alpha}\langle z, u_j\rangle v_j \\ &= \sum_{j\in J} \frac{\sigma_j^2}{\sigma_j^2 + \alpha}\frac{\langle z, u_j\rangle}{\sigma_j} v_j = \sum_{j\in J} w_j(\alpha)\frac{\langle z, u_j\rangle}{\sigma_j} v_j \end{aligned}$$

where $(w_j(\alpha))_{j\in J}$ is the Tikhonov filter. Since f_α is a strictly convex functional, it is easy to check that x_α is indeed the unique minimizer. Hence using the Tikhonov filter to obtain a regularized solution is equivalent to minimization of the Tikhonov functional f_α.

Whatever the spectral filter, the suitable choice of the regularization parameter is important to the success of the filtering. If α is too large, the singular components $\langle z, u_j\rangle$ are partially lost and with them the information about the solution. On the other hand, if α becomes too small, one obtains excessive amplification of error through small singular values. This will be clearly seen in the numerical results in Chapter 5.

Error Analysis

Let $K^\dagger z$ be the least squares solution of minimum norm to $Kx = z$. We will refer to $Px_{true} = K^\dagger K x_{true}$ as the projected true solution. Let x_α defined in (4.4) be a regularized solution to $Kx = z$, where $z = Kx_{true} + \epsilon$. Then the regularized solution error is defined by

$$e_\alpha := x_\alpha - Px_{true} = \sum_{j \in J} w_j(\alpha) \frac{\langle Kx_{true} + \epsilon, u_j \rangle}{\sigma_j} v_j - \sum_{j \in J} \langle x_{true}, v_j \rangle v_j$$

$$= \sum_{j \in J} w_j(\alpha) \frac{\langle Kx_{true}, u_j \rangle}{\sigma_j} v_j - \sum_{j \in J} \langle x_{true}, v_j \rangle v_j + \sum_{j \in J} w_j(\alpha) \frac{\langle \epsilon, u_j \rangle}{\sigma_j} v_j .$$

Since

$$\langle Kx_{true}, u_j \rangle = \langle x_{true}, K^* u_j \rangle = \sigma_j \langle x_{true}, v_j \rangle,$$

we obtain

$$e_\alpha = \sum_{j \in J} [w_j(\alpha) - 1] \langle x_{true}, v_j \rangle v_j + \sum_{j \in J} \frac{w_j(\alpha)}{\sigma_j} \langle \epsilon, u_j \rangle v_j . \tag{4.7}$$

The first component of this error is due to filtering. With α decreasing the weights $w_j(\alpha)$ tend to 1 for each j and this component approaches zero. The second component is caused by noise in the data. Its norm may become prohibitively large when α decreases to zero. Again one can see that the proper choice of the regularization parameter α is important.

Statistical Distribution of Regularized Solution Error

Consider the regularized solution error (4.7). Using our statistical model (see Chapter 3), the error can be expressed in the following way (recall the meaning of

V and U from (2.25)):

$$\begin{aligned} e_\alpha &= \sum_{j\in J}[w_j(\alpha)-1][V^*x]_j v_j + \sum_{j\in J}\frac{w_j(\alpha)}{\sigma_j}[U^*\epsilon]_j v_j \\ &= \sum_{j\in J}[\operatorname{diag}\{w_j(\alpha)-1\}V^*B_x\xi]_j v_j + \sum_{j\in J}[\operatorname{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\}U^*B_\epsilon\eta]_j v_j . \end{aligned}$$

Denoting

$$\begin{aligned} A_1 &:= \operatorname{diag}\{w_j(\alpha)-1\}V^*B_x \\ A_2 &:= \operatorname{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\}U^*B_\epsilon \end{aligned} \tag{4.8}$$

we obtain the pointwise evaluation of the regularized solution error:

$$\begin{aligned} e_\alpha(t) &:= \sum_{j\in J}[A_1\xi + A_2\eta]_j v_j(t) \\ &= \sum_{j\in J}\left[\begin{pmatrix} A_1 & A_2 \end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right]_j v_j(t) \\ &= \mathbf{v}^*(t)\begin{pmatrix} A_1 & A_2 \end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}, \end{aligned} \tag{4.9}$$

where $\mathbf{v}(t) = \begin{pmatrix} v_1(t)\\ \vdots \\ v_n(t)\end{pmatrix}$. The regularized solution error is thus a Gaussian stochastic process. For any time $t \in T$, $e_\alpha(t)$ is a normally distributed random variable. Its expected value is

$$\mathcal{E}(e_\alpha(t)) = \mathbf{v}^*(t)\begin{pmatrix} A_1 & A_2 \end{pmatrix}\mathcal{E}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = 0 \tag{4.10}$$

and its variance is

$$\begin{aligned} \operatorname{Var}(e_\alpha(t)) &= \mathcal{E}({e_\alpha}^2(t)) \\ &= \mathcal{E}\{\mathbf{v}^*(t)\begin{pmatrix} A_1 & A_2 \end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\begin{pmatrix}\xi^* & \eta^*\end{pmatrix}\begin{pmatrix}A_1^*\\ A_2^*\end{pmatrix}\mathbf{v}(t)\} \\ &= \mathbf{v}^*(t)\begin{pmatrix} A_1 & A_2 \end{pmatrix}\mathcal{E}\{\begin{pmatrix}\xi\xi^* & \xi\eta^*\\ \eta\xi^* & \eta\eta^*\end{pmatrix}\}\begin{pmatrix}A_1^*\\ A_2^*\end{pmatrix}\mathbf{v}(t). \end{aligned}$$

Because $\mathcal{E}(\xi\xi^*) = I$ and $\mathcal{E}(\xi\eta^*) = 0$ (see (3.18)), we obtain

$$\text{Var}(e_\alpha(t)) = \mathbf{v}^*(t)\begin{pmatrix} A_1 & A_2 \end{pmatrix}\begin{pmatrix} A_1^* \\ A_2^* \end{pmatrix}\mathbf{v}(t). \tag{4.11}$$

A special case of (4.11) will be needed in Chapter 5. Let $B_\epsilon = \sigma I$, $B_x = \mathcal{F}^{-1}\text{diag}\{d_j\}\mathcal{F}$ where $\mathbf{d} \in R^n$. From now on, we drop the subscript from the discrete Fourier operator $\mathcal{F}_d$. $\mathcal{F}$, $\mathcal{F}^{-1}$ are matrices associated with the FFT and IFFT, respectively (see (2.12)). Also, let $v_j(t) = \frac{1}{\sqrt{n}}e^{2\pi ijt}$. Then using (2.30) and (4.8) we obtain:

$$\begin{aligned}
A_1A_1^* &= \text{diag}\{w_j(\alpha)-1\}V^*B_xB_x{}^*V\,\text{diag}\{w_j(\alpha)-1\} \\
&= \text{diag}\{w_j(\alpha)-1\}V^*\mathcal{F}^{-1}\text{diag}\{d_j\}\mathcal{F}\mathcal{F}^*\,\text{diag}\{d_j\}(\mathcal{F}^{-1})^*V\,\text{diag}\{w_j(\alpha)-1\} \\
&= \text{diag}\{w_j(\alpha)-1\}V^*\frac{1}{\sqrt{n}}V\,\text{diag}\{d_j\}n\text{diag}\{d_j\}\frac{1}{\sqrt{n}}V^*V\,\text{diag}\{w_j(\alpha)-1\} \\
&= \text{diag}\{[(w_j(\alpha)-1)d_j]^2\}.
\end{aligned}$$

Similarly

$$\begin{aligned}
A_2A_2^* &= \text{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\}U^*B_\epsilon B_\epsilon{}^*U\,\text{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\} \\
&= \text{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\}U^*\sigma^2U\,\text{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\} \\
&= \sigma^2\,\text{diag}\left\{\left[\frac{w_j(\alpha)}{\sigma_j}\right]^2\right\}.
\end{aligned}$$

Hence (4.11) becomes

$$\text{Var}(e_\alpha(t)) = \sum_{j\in J}\frac{1}{\sqrt{n}}e^{-2\pi ijt}\left\{[(w_j(\alpha)-1)d_j]^2 + \sigma^2\left[\frac{w_j(\alpha)}{\sigma_j}\right]^2\right\}\frac{1}{\sqrt{n}}e^{2\pi ijt},$$

that is,

$$\text{Var}(e_\alpha(t)) = \frac{1}{n}\sum_{j\in J}[(w_j(\alpha)-1)d_j]^2 + \frac{\sigma^2}{n}\sum_{j\in J}\left[\frac{w_j(\alpha)}{\sigma_j}\right]^2. \tag{4.12}$$

Another random variable which is of interest is the square of the norm of the regularized solution error. Using (4.8) we have

$$
\begin{aligned}
S_\alpha := \|e_\alpha\|^2 &= \sum_{j\in J}[A_1\xi + A_2\eta]_j^2 \\
&= \|A_1\xi\|^2 + \|A_2\eta\|^2 + 2\xi^* A_1^* A_2\eta \\
&= (\xi^*, \eta^*)\, Q \begin{pmatrix} \xi \\ \eta \end{pmatrix},
\end{aligned} \tag{4.13}
$$

where

$$
Q = \begin{pmatrix} A_1^*A_1 & A_1^*A_2 \\ A_2^*A_1 & A_2^*A_2 \end{pmatrix} = \begin{pmatrix} A_1^* \\ A_2^* \end{pmatrix} \begin{pmatrix} A_1 & A_2 \end{pmatrix}. \tag{4.14}
$$

Denoting $\omega := \begin{pmatrix} \xi \\ \eta \end{pmatrix}$ we have that $\|e_\alpha\|^2 = \omega^* Q \omega$, where $\omega \sim \mathcal{N}(0, I)$. From (3.12) it follows that the random variable S_α has the distribution which is a linear combination of independent chi-square random variables. From (3.13) and the spectrum of Q we can compute the CDF of S_α (see Chapter 5).

Prediction Bands

An approximate smoothed solution x_α does not, in itself, provide information about its accuracy. To quantify accuracy we wish to derive an interval (at each point where the solution is computed) the endpoints of which are random variables that include the true value of the solution between them with probability near one, for example 0.95. This will be done separately for both random variables describing the regularized solution error:

(i) $e_\alpha(t) := (x_\alpha - Px_{true})(t), \qquad t \in T.$

(ii) $S_\alpha := \|e_\alpha\|^2.$

Consider $e_\alpha(t)$, normally distributed for every $t \in T$. Denoting

$$\mathcal{E}(e_\alpha(t)) =: \mu(t) \quad \text{and} \quad \mathrm{Var}(e_\alpha(t)) =: b(t)$$

we have (see (3.9))

$$P(|e_\alpha(t)| \le \mathrm{TOL}) = P\left(-\frac{\mathrm{TOL}-\mu(t)}{[b(t)]^{1/2}} \le Z \le \frac{\mathrm{TOL}-\mu(t)}{[b(t)]^{1/2}}\right)$$

where $Z \sim \mathcal{N}(0,1)$. Hence the value of TOL can be determined from the condition

$$P\left(-\frac{\mathrm{TOL}-\mu(t)}{[b(t)]^{1/2}} \le Z \le \frac{\mathrm{TOL}-\mu(t)}{[b(t)]^{1/2}}\right) = \gamma, \quad \gamma = 0.95.$$

Plotting the band of the width $2\times$TOL, centered at the approximate solution allows us to claim that, for each point t, with probability 0.95, the true solution, $X(t)$, will be contained within this band.

Let us now analyze, in turn, S_α. In order to know its CDF, we need the eigenvalues of the matrix Q (see (4.14)). For the numerical case considered in Chapter 5 (that is, a convolution kernel, $B_x = \mathcal{F}^{-1} D_x \mathcal{F}$, $B_\epsilon = \sigma I$), these can be computed efficiently as follows.

First using (2.30) and (4.8) we compute Q. Denote for brevity:

$$D_1 := \mathrm{diag}\{w_j(\alpha) - 1\},$$

$$D_2 := \mathrm{diag}\left\{\frac{w_j(\alpha)}{\sigma_j}\right\},$$

$$D_x := \mathrm{diag}\left\{\frac{1}{j^q}\right\}.$$

Then (see (4.8))

$$\begin{aligned} A_1^* A_1 &= (D_1 V^* B_x)^* (D_1 V^* B_x) = {B_x}^* V D_1^2 V^* B_x \\ &= (\mathcal{F}^{-1} D_x \mathcal{F})^* V D_1^2 V^* \mathcal{F}^{-1} D_x \mathcal{F} = \mathcal{F}^* D_x (\mathcal{F}^{-1})^* V D_1^2 V^* \mathcal{F}^{-1} D_x \mathcal{F} \\ &= \sqrt{n} U^* D_x \frac{1}{\sqrt{n}} V^* V D_1^2 V^* \frac{1}{\sqrt{n}} V D_x \sqrt{n} U = U^* D_x D_1^2 D_x U \\ &=: U^* D_{1x}^2 U \qquad \text{where } D_{1x} := D_x D_1. \end{aligned}$$

Similar computations yield

$$A_1^* A_2 = U^* D_{1x} D_{2x} U^* \qquad \text{where } D_{2x} := \sigma D_2,$$

$$A_2^* A_2 = U D_{2x}^2 U^*.$$

Therefore

$$\begin{aligned} Q &= \begin{pmatrix} U^* D_{1x}^2 U & U^* D_{1x} D_{2x} U^* \\ U D_{2x} D_{1x} U & U D_{2x}^2 U^* \end{pmatrix} \\ &= \begin{pmatrix} U^* & 0 \\ 0 & U \end{pmatrix} \begin{pmatrix} D_{1x}^2 & D_{1x} D_{2x} \\ D_{2x} D_{1x} & D_{2x}^2 \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix}. \end{aligned}$$

Since $\begin{pmatrix} U^* & 0 \\ 0 & U \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} = I$, the matrices $\begin{pmatrix} D_{1x}^2 & D_{1x} D_{2x} \\ D_{2x} D_{1x} & D_{2x}^2 \end{pmatrix}$ and Q are similar, and therefore have identical eigenvalues. Observe that (4.14) together with the statement

$$\text{rank}(AB) \leq \min\{\text{rank}A, \text{rank}B\}$$

show that (at least) n eigenvalues of Q are equal to zero. For the remaining n eigenvalues the following relationship holds:

$$\lambda_j = \mu_j^2 + \nu_j^2, \qquad \text{for } j = 1, 2, \cdots, n$$

where λ_j is an eigenvalue of Q, μ_j^2 is the jth eigenvalue of D_{1x}^2 and ν_j^2 is the jth eigenvalue of D_{2x}^2. To prove this, let us find an eigenvector associated with λ_1. It is easy to verify that

$$\begin{pmatrix} D_{1x}^2 & D_{1x} D_{2x} \\ D_{2x} D_{1x} & D_{2x}^2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \nu_1/\mu_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \left(\begin{array}{ccc|ccc} \mu_1^2 & \cdots & 0 & \mu_1\nu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_n^2 & 0 & \cdots & \mu_n\nu_n \\ - & - & - & - & - & - \\ \nu_1\mu_1 & \cdots & 0 & \nu_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \nu_n\mu_n & 0 & \cdots & \nu_n^2 \end{array}\right) \left(\begin{array}{c} 1 \\ 0 \\ \vdots \\ - \\ \nu_1/\mu_1 \\ \vdots \\ 0 \end{array}\right) = (\mu_1^2+\nu_1^2) \left(\begin{array}{c} 1 \\ 0 \\ \vdots \\ - \\ \nu_1/\mu_1 \\ \vdots \\ 0 \end{array}\right)$$

where ν_1/μ_1 is the $(n+1)$st component of the eigenvector. Note that $j = 1$ was chosen for notational convenience only.

Once the eigenvalues of Q are available we have determined CDF of S_α. Again an equation for TOL can be set up

$$F(S_\alpha \leq \text{TOL}) = \gamma.$$

Knowing TOL we determine the prediction intervals according to numerical procedure described in Chapter 5.

CHAPTER 5

NUMERICAL RESULTS

The numerical results showing the behavior of filtered and unfiltered solutions in both the time and the frequency domain are first presented. Then prediction bands based on the random variables $e_\alpha(t)$ and S_α are plotted for various test cases. Error indicators and the dependence of filtered solutions on the regularization parameter α are discussed.

All computations were performed on an IBM PC. The integral equation $Kx = y$ with a convolution kernel has a specific singular system (2.29) which allows us to apply the Fast Fourier Transform (see (2.12)) to obtain the results very efficiently even for high dimensional subspaces (n=512).

The Effect of Filtering

Consider an integral equation with a convolution kernel

$$y(s) = (Kx)(s) = \int_0^1 k(s-t)x(t)\,dt$$

which can also be written as $y = k * x$. Here $K : H_0^p(0,1) \to L^2(0,1)$. Therefore $\hat{y} = \hat{k}\hat{x}$ (see (2.11)). Synthetic data $\hat{y}$ from the deterministic true solution:

$$x(t) = \begin{cases} 10t^2, & 0 \le t < 0.25; \\ \frac{0.02885}{t-0.2038}, & 0.25 \le t < 0.3; \\ 1.1149e^{-20(t-0.55)^2} - 0.019429, & 0.3 \le t \le 1; \end{cases}$$

is generated. To the data $\hat{y}$, the Fourier transform of a pseudorandom noise vector $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, is added. The standard deviation σ is related to the signal-to-noise