Conjugate Gradient in Hilbert Spaces

Ivo Panayotov

Department of Mathematics and Statistics,
McGill University, Montréal
Québec, Canada

August, 2004

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements of the degree of
Master of Science

Your file Votre référence
ISBN: 0-494-12517-9
Our file Notre référence
ISBN: 0-494-12517-9

Abstract

In this thesis, we examine the *Conjugate Gradient* algorithm for solving self-adjoint positive definite linear systems in $\mathbb{C}^n$. We generalize the algorithm by proving a convergence result of *Conjugate Gradient* for self-adjoint positive definite operators in an arbitrary Hilbert space H. Then, we use the Maple software for symbolic manipulation to implement a general version of *Conjugate Gradient* and to demonstrate, by examples, that the algorithm can be used directly to solve problems in Hilbert spaces other than $\mathbb{C}^n$.

Résumé

Dans cette thèse, nous examinons l'algorithme de *Gradient Conjugué* pour résoudre des systèmes d'équations linéaires auto-adjoints definis positifs dans $\mathbb{C}^n$. Nous généralisons l'algorithme en prouvant un résultat de convergence de *Gradient Conjugué* pour les opérateurs auto-adjoints definis positifs dans un espace d'Hilbert arbitraire. Ensuite, nous employons le logiciel de manipulation symbolique *Maple* afin d'implémenter une version générale de *Gradient Conjugué.* Nous démontrons ainsi, par des examples, que l'algorithme peut être utilisé directement pour résoudre des problèmes dans des espaces d'Hilbert autres que $\mathbb{C}^n$.

Acknowledgments

First, I wish to express my deepest gratitude to my supervisor professor Paul Tupper for his guidance, and invaluable comments and suggestions throughout the research, and for bearing with me and being at my disposal, throughout the day, during my last moments of stress before the submission. Also special thanks to professor Martin Gander who got interested in the project and suggested me to work on Conjugate Gradient rather than GMRES. Without him I would still be stuck doing GMRES, which as it turned out is considerably harder. Special thanks also goes to Eugene Kritchevski, always one step ahead in the Spectral theory, for discussing with me my own problems and giving me many useful insights. Thanks to Paul Koosis who had patience and didn't chase me immediately from his office for asking obvious questions. Finally I would like to acknowledge Marco Bertola, from the university of Concordia, who gave a superb course on Maple which I audited. Without his course I would not have done this project.

Table of Contents

Abstract i

Résumé iii

Acknowledgments v

Introduction 1

1 Conjugate Gradient in $\mathbb{C}^n$ 5

1.1 Introduction to Krylov Subspaces 5

1.2 Self-Adjoint Positive Definite Matrices 7

1.3 Conjugate Gradient 9

2 Hilbert Space and Linear Operator Theory 19

2.1 Inner-Product and Normed Spaces (Review) 20

2.2 Metric Spaces (Review) 23

2.3 Banach and Hilbert Spaces 23

2.4 Linear Operators on Hilbert Spaces 25

2.5 Spectral Theory of Linear Operators 30

3 Conjugate Gradient in Hilbert Spaces 33

3.1 Spectral Properties of Self-Adjoint Positive Linear Operators 34

3.2 Convergence of CG in Hilbert Spaces 42
3.3 Examples. Integral Equations. 46
3.4 CG in practice 66

Conclusion **71**

A Conjugate Gradient for Self-Adjoint Positive Definite Operators **73**

B Maple Code Used in the Examples **77**

Introduction

In this thesis, we consider the equation $Lu = f$ where L is a linear operator, u is an unknown real or complex-valued function over a domain $\mathcal{D}$, and f is a known function. Solving such equations is of a high interest among mathematicians and scientists, since many differential and integral equations can be written in the above manner. Closed form solutions to such equations do not always exist and one usually uses a numerical method to approximate the solutions. The main procedure for doing this consists of two steps:

1. Discretizing the domain $\mathcal{D}$ and the linear equation $Lu = f$ in order to obtain a linear system $Ax = b$ over $\mathcal{D}_d$, where $\mathcal{D}_d$ consists of finitely many sample points from $\mathcal{D}$. Here, A is a matrix, and x and b are vectors representing the discrete versions of u and f respectively over $\mathcal{D}_d$.

2. Solving the system $Ax = b$.

In performing the above steps, one hopes that the solution vector x, of the discrete system, will be sufficiently close to the solution u, of the original system, over $\mathcal{D}_d$. Usually, the discretization of $\mathcal{D}$ in step 1 needs to be sufficiently refined for the matrix problem $Ax = b$ to give a reasonable approximation to $Lu = f$. In most real problems, A is so large that computing the exact solution of the discrete system via a direct method, such as Gaussian elimination, is not feasible. Moreover, even if one could solve $Ax = b$ exactly, the vector x will be only an approximation to the function

u over $\mathcal{D}_d$. For these reasons one usually settles for a good approximate solution of the discrete system.

Throughout the years, a considerable number of iterative methods have been developed in order to efficiently approximate the solutions of large linear systems $Ax = b$. This paper examines one such method called *Conjugate Gradient (CG)*, which is a Krylov subspace method (the terminology will be explained later), and can be only applied to self-adjoint (Hermitian) positive definite matrices A. Due to its excellent performance in solving real-life problems, CG is currently the most popular method for solving such systems.

Our objective in this thesis is, instead of applying CG to discrete systems, to apply it directly to problems $Lu = f$ avoiding the discretization step completely. Thus, the two-step procedure for solving $Lu = f$ described earlier will become a one-step procedure, namely:

1. Apply CG directly to $Lu = f$.

For this to be feasible, L needs to satisfy similar properties to those required of A, but this will be discussed in due time.

Our thesis is outlined as follows. In Chapter 1, we examine the CG algorithm for solving discrete linear systems $Ax = b$ where A is self-adjoint positive definite. We describe the method in some detail mainly focusing on its convergence properties. We also give an example of how to discretize a linear equation $Lu = f$ in order to obtain a system $Ax = b$. Then, in Chapter 2, we introduce the notion of a *Hilbert space*, generalizing $\mathbb{C}^n$, and extend some key results from matrix theory to linear operators on a Hilbert space. Finally, in Chapter 3, we define the concept of a self-adjoint positive definite linear operator on a Hilbert space and prove a convergence result for CG

applied to this type of operators. We end the chapter by solving several equations $Lu = f$ using a direct application of *CG*, implemented in the *Maple* software for symbolic computation. The important part of our work is Chapter 3. An advanced reader, who understands *CG* well and has a good grasp of operator theory, can easily omit chapters 1 and 2 which are primarily intended as prerequisites for Chapter 3.

We now shortly describe how this project came about. Having a general interest in matrix iterations, as well as some knowledge of *Maple*, our basic interest was to determine whether it is possible to use matrix iterative schemes to solve continuous linear equations symbolically, by applying the scheme directly to a given equation. Our research started with a Krylov subspace method called *Generalized Minimal Residuals (GMRES)* which is more general than *CG* in that it does not require the matrix A to be self-adjoint positive definite. Having studied *GMRES* for finite dimensional linear systems, both in the case where the matrix A is invertible ([25] and [27, pp.266–275]) and in the case where A is non-invertible ([6] and [17]), we implemented a *Maple* procedure which successfully solved some simple linear problems symbolically. We then undertook a more in-depth study of *GMRES* by examining, in [8, 9, 12, 21], some convergence-rate results for a family $A_\gamma x = b$ of linear equations with the spectrum of the A_γ lying in a common domain. Such families usually arise in the successive refinement of the mesh for a particular continuous problem and the resulting convergence-rate estimate is common to the whole family of equations, that is, the estimate is mesh-independent. In [8, 9] the authors also analyzed the convergence behavior of *GMRES* in a Hilbert space for operators of the type $I - K$ where I denotes the identity operator, K is a compact operator, and $I - K$ is invertible. Later, we turned our attention to the more popular *CG* algorithm which can, however, only be applied under the restriction that A is self-adjoint positive definite. Analogously, we studied *CG* in the finite dimensional case ([3, pp.459–480, pp.558–593], [4], [13,

pp.520–542], [26] and [27, pp.293–302]), as well as in a Hilbert space ([5] and [10, pp.114–136]).

There are two remarkable features about this thesis which are worth pointing out. First, we have given a proof of Lemma 3.1, used to prove convergence of *CG*, which we believe is new. Although the convergence of *CG* in a Hilbert space has been proven in [10, pp.117–121], its proof uses the Spectral Theorem. Our proof has the merit of being more elementary since it does not require the use of the Spectral Theorem, but is only based on the existence of a positive square root for a self-adjoint positive definite operator, which can be proven by elementary means. Second and more important, we have implemented, in *Maple*, a symbolic version of *CG* allowing us to solve self-adjoint positive definite equations directly as can be seen from the examples in Chapter 3. There, we have solved several integral equations with *CG* by symbolic manipulation.

Chapter 1

Conjugate Gradient in $\mathbb{C}^n$

The aim of this chapter is to present the basic ideas behind the *CG* algorithm for solving discrete linear systems. We take the point of view that *CG* is a Krylov subspace method and describe *what CG* does rather than exactly *how* it does it. The reader, who has never encountered *CG* before and wishes to understand the exact mechanics behind the algorithm, is urged to read the excellent paper by Jonathan Shewchuk [26]. Other important, but denser and more advanced readings, include [13, pp.520–541], [27, pp.293–302], [11, pp.307–319] and [14, pp.33-37, pp. 49–54]

1.1 Introduction to Krylov Subspaces

Consider the system $Ax = b$ where $A \in \mathbb{C}^{n\times n}$, $b \in \mathbb{C}^n$ is known and $x \in \mathbb{C}^n$ is unknown. We will further assume that the matrix A is nonsingular ensuring that the system has a solution. The idea of Krylov methods is to construct the sequence of vectors $b, Ab, A^2b, \ldots$, and to look for the "best" approximation of the solution vector $\hat{x} = A^{-1}b$ inside each of the subspaces

$$K_0 = \{0\},$$

$$K_1 = span\{b\},$$

$$K_2 = span\{b, Ab\},$$

$$K_3 = span\{b, Ab, A^2b\},$$

$$\vdots$$

The exact meaning of the term "best" varies according to the specific Krylov method used to solve the system. We will shortly make the concept precise in the case of *CG*. The subspaces of $\mathbb{C}^n$ constructed above are called *Krylov subspaces* corresponding to A and b. Sometimes they are denoted by $K_j(A, b)$; here, we simply write K_j as it is clear from the context to which matrix and vector they are referred. The sequence of Krylov subspaces of $\mathbb{C}^n$ is finite and nested, and has the property that one of its members, some K_j, contains the solution vector $\hat{x}$ as the following proposition asserts.

Proposition 1.1. *If A is invertible, the solution $\hat{x}$ of the system $Ax = b$ is a member of some Krylov subspace $K_j \subseteq \mathbb{R}^n$ constructed in the above manner.*

Proof. Let the minimal polynomial of A be given by

$$p(t) = a_0 + a_1 t + \ldots + a_j t^j$$

where $j > 0$ and $a_j \neq 0$. Since $0 = p(A)$, we have

$$0 = p(A)b = a_0 b + a_1 Ab + \ldots + a_j A^j b,$$

hence

$$a_0 b = -a_1 Ab - a_2 A^2 b - \ldots - a_j A^j b.$$

Notice that $a_0 \neq 0$ otherwise $A^{-1}p(A)$ would be a polynomial in A also equal to 0 contradicting the minimality of p. Dividing by a_0 and multiplying by A^{-1} on both sides yields

$$\hat{x} = A^{-1}b = -\frac{a_1}{a_0}b - \frac{a_2}{a_0}Ab - \ldots - \frac{a_j}{a_0}A^{j-1}b \in K_j$$

as required. □

1.2 Self-Adjoint Positive Definite Matrices

Before we go on, we need to put a precise meaning to the concept of a self-adjoint positive definite matrix, which will be used throughout the chapter. This concept is standard and the reader is probably well aware of it, but we define it for completeness. We also state some fundamental results about such matrices on which the *CG* algorithm is build. They can be found in the Linear Algebra books [2, 20, 22]. Since these results are very basic, we will use them later on without explicitly referring to them.

Definition 1.1. Let $A \in \mathbb{C}^{n\times n}$. The *adjoint* of A, denoted by A^*, is the conjugate transpose of A, that is $A^*_{ij} = \overline{A_{ji}}$ for all i, j.

Definition 1.2. Let $A \in \mathbb{C}^{n\times n}$. Then A is called *self-adjoint* (or *Hermitian*) if $A = A^*$.

When A is real, A^* is just A^T. In that case, the concept of self-adjointness reduces to *symmetry*. Another important observation concerning a self-adjoint matrix A is that $\langle Ax, y\rangle = \langle x, Ay\rangle = x^*Ay$ for all $x, y \in \mathbb{C}^n$ where $\langle\cdot,\cdot\rangle$ denotes the standard inner product on that space.

Proposition 1.2. *If $A \in \mathbb{C}^{n\times n}$ is self-adjoint and $x \in \mathbb{C}^n$, then x^*Ax is real.*

Proof. Notice that $x^*Ax \in \mathbb{C}$. To prove the proposition, it is therefore sufficient to show that $\overline{x^*Ax} = x^*Ax$. Since $A^* = A$, we have

$$\overline{x^*Ax} = \left(\overline{x^*Ax}\right)^T = (x^*Ax)^* = x^*A^*x = x^*Ax.$$

□

Definition 1.3. Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. If $x^*Ax > 0$ for all non-zero vectors $x \in \mathbb{C}^n$ then A is called *positive definite*.

Proposition 1.3. *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint positive definite. Then A can be used to define an inner product on $\mathbb{C}^n$ by*

$$\langle x, y\rangle_A := \langle Ax, y\rangle = x^*Ay. \tag{1.1}$$

Proof. We omit it since later we prove a similar but more general result (Proposition 3.1). Also see [20, p. 276]. □

Definition 1.4. Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint positive definite. Then for any $x \in \mathbb{C}^n$ we define the A-norm of x by

$$\|x\|_A := \langle Ax, x\rangle_A^{\frac{1}{2}}. \tag{1.2}$$

Proposition 1.4. *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint positive definite. Then all eigenvalues of A are strictly positive real numbers.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A with corresponding eigenvector x. Then $\lambda x = Ax$. Hence, $x^*\lambda x = x^*Ax$ implying $\lambda = \frac{x^*Ax}{x^*x}$ which is a real and strictly positive. □

Proposition 1.5. *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint positive definite. Then A is invertible.*

Proof. From the previous proposition, it follows that zero is not an eigenvalue of A. Hence A is invertible. □

Thus, for a self-adjoint positive definite matrix A, the system $Ax = b$ has a unique solution which, as discussed in the previous section, lies in some Krylov subspace K_j. We now state one last definition, after which we are ready to introduce the *CG* algorithm for solving this kind of linear systems.

Definition 1.5. Suppose x_j approximates the solution $\hat{x}$ of the equation $Ax = b$. We define the *error* of approximation as $e_j = x_j - \hat{x}$.

1.3 Conjugate Gradient

We now turn our attention to the Conjugate Gradient algorithm. Given a matrix $A \in \mathbb{C}^{n\times n}$, which is self-adjoint positive definite, and a vector $b \in \mathbb{C}^n$, the algorithm approximates the unique solution of $Ax = b$ by:

- Constructing, at step j, the vectors $b, Ab, A^2b, \ldots A^{j-1}b$ which span K_j
- Finding $x_j \in K_j$ which minimizes $\|e_j\|_A$.

Minimizing $\|e_j\|_A$ is a good way of controlling the error since $e_j \to 0$ componentwise if and only if $\|e_j\|_A \to 0$, and $e_j = 0$ if and only if $\|e_j\|_A = 0$. This is just a consequence of the fact that on a finite dimensional vector space over $\mathbb{C}$ all norms are equivalent; in particular, the A-norm is equivalent to the ∞-norm. Of all possible norms, why minimize e_j in the A-norm? There are two reasons for this. The first one is efficiency. It turns out that minimizing the A-norm of the error in K_j allows the j^{th} iterate to be written as $x_j = x_{j-1} + \alpha A^{j-1}b$ where α is a scalar [13, pp. 522]. In fact, α is real valued because it represents the step length of the improvement at step j in the direction of $A^{j-1}b$. The j-dimensional optimization problem of finding x_j in K_j is reduced, by using the information from the previous iterates, to a one-dimensional optimization problem, namely, finding the value of α. The second reason is that finding $x_j \in K_j$ for which $\|e_j\|_A$ is minimal does not require any information about $\hat{x}$ other than $A\hat{x} = b$. This can be easily seen by writing

$$\begin{aligned}
\|e_j\|_A^2 &= \langle Ae_j, e_j\rangle \\
&= \langle A(x_j - \hat{x}), x_j - \hat{x}\rangle \\
&= \langle Ax_j, x_j\rangle - \langle Ax_j, \hat{x}\rangle - \langle A\hat{x}, x_j\rangle + \langle A\hat{x}, \hat{x}\rangle \\
&= \langle Ax_j, x_j\rangle - \langle x_j, b\rangle - \langle b, x_j\rangle + \langle A\hat{x}, \hat{x}\rangle\,.
\end{aligned}$$

Notice that the term $\langle A\hat{x}, \hat{x}\rangle$ in the last expression is unknown but constant. Hence, minimizing $\|e_j\|_A^2$, and consequently $\|e_j\|_A$, with respect to x_j is equivalent to mini-

mizing $\langle Ax_j, x_j\rangle - \langle x_j, b\rangle - \langle Ab, x_j\rangle$; this is done without knowledge of the exact value of $\hat{x}$.

Now, since the Krylov subspaces constructed by CG are nested, it follows that

$$\|e_0\|_A \geq \|e_1\|_A \geq \|e_2\|_A \geq \dots.$$

Moreover we have the following convergence result.

Proposition 1.6. *In the absence of rounding errors, the* CG *algorithm converges to the exact solution of $Ax = b$ in at most k steps, where k is the degree of the minimal polynomial of A.*

Proof. This proposition is an immediate consequence from the statement and proof of Proposition 1.1. □

We now give another, much more complicated proof of the convergence of CG by estimating the rate of convergence of the sequence $\|e_j\|_A$. This proof is important, as it will give us insight in our later discussion about function spaces which are infinite-dimensional. There, we cannot hope for CG to find the exact solution at any stage. Before we state and prove the main theorem, we will need the following definitions and technical lemmas. The discussion here follows closely the one in Trefethen and Bau [27, pp. 293–302]. In particular, Lemma 1.2 and Theorem 1.8 are modified and expanded versions of the ones given in that text.

Definition 1.6. Let $A \in \mathbb{C}^{n\times n}$. We define the *spectrum* of A, denoted by $\sigma(A)$, as the set of all eigenvalues of A.

Definition 1.7. Define $P_j := \{$*real polynomials of degree at most j with* $p(0) = 1\}$.

Lemma 1.1. *The error at the j^{th} iteration of CG can be expressed as $e_j = p_j(A)e_0$, for some $p_j \in P_j$. Hence, at step j CG minimizes $\|p_j(A)e_0\|_A$ over P_j.*

Proof. Since the starting approximation of CG is inside $K_0 = \{0\}$, then the starting error is just $e_0 = 0 - \hat{x} = -\hat{x}$. The approximation at the j^{th} iteration is given by

$$x_j = \alpha_1 b + \alpha_2 Ab + \ldots + \alpha_j A^{j-1} b = \alpha_1 A\hat{x} + \alpha_2 A^2\hat{x} + \ldots + \alpha_j A^j \hat{x} = q(A)\hat{x}$$

for some polynomial q with $q(0) = 0$. Therefore the j^{th} error is just

$$e_j = x_j - \hat{x} = q(A)\hat{x} - \hat{x} = (q(A) - I)\hat{x} = (I - q(A))e_0 = p_j(A)e_0$$

for a polynomial p_j satisfying $p_j(0) = 1$. The second part of the lemma follows. Notice that the α_j are real valued by a remark on page 9. □

Lemma 1.2. *If at step j the CG iteration has not already converged, we have*

$$\|e_j\|_A = \inf_{p \in P_j} \|p(A)e_0\|_A \leq \left(\inf_{p \in P_j} \max_{\lambda \in \sigma(A)} |p(\lambda)| \right) \cdot \|e_0\|_A \tag{1.3}$$

Proof. The equality is just the result stated in the previous lemma. We now prove the inequality. Since A is self-adjoint, we can construct an orthonormal basis $\{v_1, v_2, \ldots, v_n\}$ for $\mathbb{C}^n$ composed of eigenvectors of A. Hence we can express e_0 in terms of this basis as

$$e_0 = \sum_{k=1}^{n} \alpha_k v_k$$

Then,

$$\begin{aligned} \|e_0\|_A^2 &= \left\langle A \sum_{k=1}^{n} \alpha_k v_k, \sum_{k=1}^{n} \alpha_k v_k \right\rangle \\ &= \left\langle \sum_{k=1}^{n} \alpha_k \lambda_k v_k, \sum_{k=1}^{n} \alpha_k v_k \right\rangle \\ &= \sum_{k=1}^{n} \alpha_k^2 \lambda_k, \end{aligned}$$

where the last equality holds by orthonormality of the v_k. Now, consider an arbitrary polynomial $p \in P_j$. Since $p(A)v_k = p(\lambda_k)v_k$ where λ_k is the eigenvalue of v_k, we can write

$$p(A)e_0 = \sum_{k=1}^{n} \alpha_k p(\lambda_k) v_k.$$

Therefore,

$$\begin{aligned}
\|p(A)e_0\|_A^2 &= \left\langle A\sum_{k=1}^{n}\alpha_k p(\lambda_k)v_k, \sum_{k=1}^{n}\alpha_k p(\lambda_k)v_k \right\rangle \\
&= \left\langle A\sum_{k=1}^{n}\alpha_k p(\lambda_k)v_k, \sum_{k=1}^{n}\alpha_k p(\lambda_k)v_k \right\rangle \\
&= \sum_{k=1}^{n}\alpha_k^2 p^2(\lambda_k)\lambda_k \\
&\leq \max_{\lambda\in\sigma(A)}|p(\lambda)|^2 \sum_{k=1}^{n}\alpha_k^2\lambda_k \\
&= \left(\max_{\lambda\in\sigma(A)}|p(\lambda)|\right)^2 \|e_0\|_A^2.
\end{aligned}$$

Taking square roots and using the fact that p is arbitrary, we deduce that

$$\inf_{p\in P_j}\|p(A)e_0\|_A \leq \left(\inf_{p\in P_j}\max_{\lambda\in\sigma(A)}|p(\lambda)|\right)\|e_0\|_A.$$

□

This lemma shows that it is possible to estimate the convergence rate of CG by finding polynomials which, as j increases, have smaller and smaller magnitude on the spectrum of A. Since we do not know a priori how the spectrum of A is distributed, it is difficult to choose the very best polynomials for this purpose. We can, however, choose polynomials which are the smallest possible (in the infinity norm) along the interval $[\lambda_{min}, \lambda_{max}]$ where λ_{min} and λ_{max} correspond respectively to the smallest and largest eigenvalues of A. These eigenvalues are both real and strictly positive because A is self-adjoint and positive definite. The polynomials in question are the scaled and shifted Chebyshev polynomials. We assume that the reader is familiar with the standard Chebyshev polynomials, so we discuss them only briefly, stating, without proof, some of their properties. These can be found in [7, pp. 507–517].

Chebyshev Polynomials

The Chebyshev polynomials are defined for each $n = 0, 1, \ldots$ by $T_n = \cos(n \arccos x)$. Written in this form, their domain is just $[-1, 1]$. However, since they can be rewritten in standard polynomial form, we can extend their domain to the whole real line. These polynomials have the following properties.

1. $T_0 = 1$, $T_1 = x$, and $T_j(x) = 2xT_{j-1}(x) - T_{j-2}(x)$ for $j > 1$.
2. T_n takes only values between -1 and 1 in the interval $[-1, 1]$.
3. T_n has n zeros, all inside $[-1, 1]$.
4. T_n takes strictly positive values for $x > 1$.
5. $\max_{[-1,1]} |T_n|$ is the smallest possible from all polynomials of degree n having the same leading coefficient as T_n.

Property five is the reason we have chosen Chebyshev polynomials in our discussion. If we wish, for example, to find the monic polynomial of degree n, whose maximum in $[-1, 1]$ is the smallest possible, we have to pick $p(x) = \alpha T_n(x)$ where α is chosen so that $p(x)$ is itself monic. We are finally ready to prove our main result estimating the convergence rate of CG. We point out that the estimate given below does not provide a sharp bound for the rate of convergence of CG. Since we are mainly interested in establishing the convergence of CG, only using rate of convergence as an argument, we have decided to prove a weaker but simpler result. The sharp result can be found in [3, pp. 566-567], [27, pp. 299-300], [13, p. 530].

Theorem 1.8. *Let CG be applied to the problem $Ax = b$, where $A \in \mathbb{C}^{n \times n}$ is a self-adjoint positive definite matrix. Also, let λ_{min} and λ_{max} denote respectively the smallest and largest eigenvalue of A. Then the errors of approximation satisfy*

$$\|e_j\|_A \leq \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right)^j \|e_0\|_A \tag{1.4}$$

Proof. First, consider the trivial case where $\lambda_{min} = \lambda_{max}$. Since A is self-adjoint, hence diagonalizable, it is similar to the scalar matrix λI. Thus $A = Q^{-1}\lambda IQ$ for some invertible matrix Q. This can only happen if $A = \lambda I$. In this case the solution to $Ax = b$, is just $\hat{x} = b/\lambda \in K_1$. Thus CG converges in one iteration and the statement of the theorem holds.

Assume now that $\lambda_{max} > \lambda_{min}$. The transformation

$$t \mapsto \frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}} - \frac{2}{\lambda_{max} - \lambda_{min}} t$$

maps the interval $[\lambda_{min}, \lambda_{max}]$ to $[-1, 1]$. This transformation inverts the orientation of the real line, but we have chosen it for convenience. Set

$$p_j(t) := \frac{T_j\left(\frac{\lambda_{max}+\lambda_{min}}{\lambda_{max}-\lambda_{min}} - \frac{2}{\lambda_{max}-\lambda_{min}} t\right)}{T_j\left(\frac{\lambda_{max}+\lambda_{min}}{\lambda_{max}-\lambda_{min}}\right)}. \tag{1.5}$$

Then we have $p_j(0) = 1$. The numerator of (1.5) is between -1 and 1 for $t \in [\lambda_{min}, \lambda_{max}]$ by Property 2 of Chebyshev polynomials. We now show that the denominator satisfies

$$T_j\left(\frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}}\right) \geq \left(\frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}}\right)^j, \tag{1.6}$$

which in particular makes p_j well defined. That it is well-defined also follows immediately from properties 3 or 4 of Chebyshev polynomials. Let

$$\gamma = \frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}}.$$

Clearly $\gamma > 1$. By Property 1,

$$T_0(\gamma) = 1 < \gamma = T_1(\gamma) < 2\gamma T_1(\gamma) - T_0(\gamma) = T_2(\gamma) < \ldots < 2\gamma T_{j-1}(\gamma) - T_{j-2}(\gamma) = T_j(\gamma)$$

so the sequence $T_j(\gamma)$ is increasing. Again using Property 1, and the fact that $T_j(\gamma)$ increases, we have:

$$T_0(\gamma) \geq 1,$$

$$T_1(\gamma) \geq \gamma,$$

$$T_2(\gamma) = 2\gamma T_1(\gamma) - T_0(\gamma) \geq \gamma T_1(\gamma) \geq \gamma^2,$$

$$\vdots$$

$$T_j(\gamma) = 2\gamma T_{j-1}(\gamma) - T_{j-2}(\gamma) \geq \gamma T_{j-1}(\gamma) \geq \gamma^j,$$

which proves equation (1.6). Finally, since $p_j(t)$ defined by (1.5) belongs to P_j, then

$$\inf_{p \in P_j} \max_{\lambda \in \sigma(A)} |p(\lambda)| \leq \max_{\lambda \in \sigma(A)} |p_j(\lambda)| \leq \max_{t \in [\lambda_{min}, \lambda_{max}]} |p_j(t)| \leq \frac{1}{T_j(\gamma)} \leq \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right)^j.$$

Lemma (1.2) now yields

$$\|e_j\|_A \leq \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right)^j \|e_0\|_A$$

as required. □

This theorem shows that CG converges and gives an upper bound on the rate of convergence. A little more careful analysis reveals that in fact,

$$T_j\left(\frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}} \right) = \frac{1}{2} \left[\left(\frac{\sqrt{\lambda_{max}} + \sqrt{\lambda_{min}}}{\sqrt{\lambda_{max}} - \sqrt{\lambda_{min}}} \right)^j + \left(\frac{\sqrt{\lambda_{max}} + \sqrt{\lambda_{min}}}{\sqrt{\lambda_{max}} - \sqrt{\lambda_{min}}} \right)^{-j} \right] \tag{1.7}$$

giving a better estimate of the rate of convergence of CG. Since our primary concern is not efficiency but convergence, the estimate given in the theorem, being much simpler, is sufficient.

Example 1.1. *Suppose we want to solve the following one-dimensional steady-state heat equation on the interval* $[0, 1]$,

$$-\triangle u = 12t^2,\ u(0) = 0,\ u(1) = 0.$$

This is a boundary value problem which can be easily solved by integration; the exact solution is given by $u(t) := t - t^4$. *We will solve it, however, by first discretizing the problem and then applying* CG *to the resulting system of linear equations. First,*

we discretize the domain $[0,1]$ *by selecting a certain number of points from it. For simplicity, we choose 5 points at equal intervals of length* $h = 0.25$. *Thus, our discrete domain is simply represented by the vector* $(0, 0.25, 0.5, 0.75, 1)$. *Now, we assume that we can write a third degree Taylor expansion of* u. *Then,*

$$u(t+h) = u(t) + hu'(t) + \frac{h^2}{2}u''(x) + \mathcal{O}(h^3),$$

$$u(t-h) = u(t) - hu'(t) + \frac{h^2}{2}u''(x) + \mathcal{O}(h^3).$$

Adding the two expressions and rearranging the result produces the formula

$$\frac{u(t-h) - 2u(t) + u(t+h)}{h^2} = u''(t) + \mathcal{O}(h).$$

Thus, the expression on the left approximates $u''(t)$, *and, as* $h \to 0$, *the error of approximation* $\to 0$. *We now let*

$$\begin{aligned} u_1 &:= u(0), \\ u_2 &:= u(0.25), \\ u_3 &:= u(0.5), \\ u_4 &:= u(0.75), \\ u_5 &:= u(1). \end{aligned}$$

Thus, in our discrete version of the problem, we replace the function $u(t)$ *by the vector* $\overline{u} = (u_i)$. *We also evaluate* $12\,t^2$ *at each point of the partition to obtain* $(0, 0.75, 3, 6.75, 12)$. *Using the above approximation for* $u''(t)$, *we are now ready to write the following system of linear equations approximating the original problem:*

$$\begin{aligned} u_1 &= 0, \\ \frac{-u_0 + 2u_1 - u_2}{h^2} &= 0.75, \\ \frac{-u_1 + 2u_2 - u_3}{h^2} &= 3, \\ \frac{-u_2 + 2u_3 - u_4}{h^2} &= 6.75, \\ u_5 &= 0. \end{aligned}$$

This system can be written in matrix form as,

$$\frac{1}{0.25^2}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.75 \\ 3 \\ 6.75 \\ 0 \end{pmatrix},$$

where the matrix above is not yet self-adjoint. However, knowing that $u_1 = u_5 = 0$, *we can rewrite the above system as*

$$\frac{1}{0.25^2}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.75 \\ 3 \\ 6.75 \\ 0 \end{pmatrix},$$

where this time the matrix is symmetric and positive definite. The last property can be verified directly from Definition 1.3 for any matrix of the above form, no matter its size. Applying CG *to the above system produces the results shown in Figure 1.1. It takes 4 iterations for the algorithm to find the exact solution of the discrete system. The points of a particular approximation are joined by (dotted) lines so it is easier to group them. The solid curve represents the true solution of the original equation whereas its closest dotted curve represents the exact solution of the discrete system at the required five points. The implementation of* CG *used here is given in Appendix A, and the Maple code used to generate the example is given in Appendix B.*

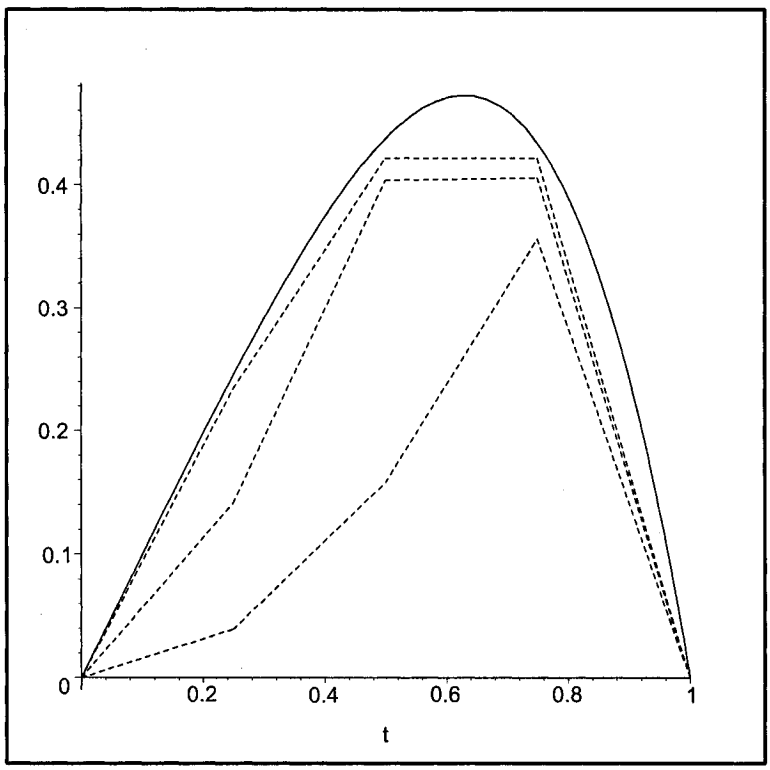

Figure 1.1: Results from *CG* for Example1.1

Chapter 2

Hilbert Space and Linear Operator Theory

In a vector space, we can add vectors and also multiply a vector by a scalar. We can also introduce the concept of a *norm* which generalizes the notion of *length* that we have from working with the usual two- and three-dimensional real vectors. This gives rise to the so called *normed vector spaces.* We can go even further and define an *inner product* of vector which generalizes the notion of the usual *scalar product* on $\mathbb{R}^2$ or $\mathbb{R}^3$. This inner product allows us to define the further concepts of *orthogonality* and *angle* between vectors. It should be noted that the inner product can also be used to define the norm itself. Vector spaces equiped with an inner product are called *inner product spaces.* If an inner product space is also *complete* with respect to the norm induced by its inner product, it is called a *Hilbert space.* These Hilbert spaces are the ones that are of interest to us because they generalize appropriately $\mathbb{R}^n$ and $\mathbb{C}^n$ for which *CG* was originally designed. In the first half of this chapter we explain more carefully the meaning of each concept used above. Then, in the second half, we define the concept of a *linear operator* on a Hilbert space and discuss some important properties of linear operators which will be used later on.

In $\mathbb{R}^n$ or $\mathbb{C}^n$, linear operators can be represented as matrices in the sense that applying an operator to an element of the vector space is equivalent to a matrix-vector multiplication. As a result, the study of properties of operators in these vectors spaces reduces to the study of the properties of their matrix representatives. In this thesis, we consider spaces of arbitrary dimension. We will see that many of the standard properties of matrices can be suitably extended to linear operators on an arbitrary Hilbert space H. The reader should be warned, however, that there are also many complications arising from the fact that H may be infinite dimensional.

We point out that the current chapter provides only a brief overview of the vast field of operator theory. Although the reader is not required to have an in-depth knowledge on the subject, it is not the aim of our work to build a complete background. For this reason, we will often state results and omit the proofs as even some basic facts have lengthy proofs and require previous background. The reader who wishes to gain further understanding of the subject should look at some of the references given in the bibliography [15, 18, 19, 24].

2.1 Inner-Product and Normed Spaces (Review)

This section, as well as the next, are review sections of fundamental linear algebra and analysis concepts which form the basic setup for our later discussion.

Definition 2.1. Let X be a vector space over $\mathbb{C}$. An *inner product* on X is a mapping $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{C}$ satisfying the following properties for all $x, y, z \in X$ and scalars a, b:

1. $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$ (linearity of the first component),

2. $\langle x, y\rangle = \overline{\langle y, x\rangle}$ (conjugate symmetry),

3. $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$. (positivity)

Definition 2.2. A vector space X with an inner product is called an *inner product space.*

Note that properties (1) and (2) imply that the inner product is conjugate linear in the second component, that is, $\langle x, ay + bz\rangle = \overline{a}\,\langle x, y\rangle + \overline{b}\,\langle y, z\rangle$. Also, when X is a real vector space, the inner product becomes linear in both components as well as symmetric. We now define the concept of a normed space.

Definition 2.3. Let X be a vector space over $\mathbb{C}$. A *norm* on X is a mapping $\|\cdot\| : X \to \mathbb{R}$ satisfying the following properties for all $x, y \in X$ and scalars k:

1. $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$,

2. $\|kx\| = |k|\|x\|$,

3. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

Definition 2.4. A vector space X with a norm is called a *normed vector space.*

An inner product on a vector space X can be used to define a norm by

$$\|x\| = \sqrt{\langle x, x\rangle}. \tag{2.1}$$

The norm defined in this way is said to be *induced* by the inner product on X. In general, not all norms are induced by inner products. In fact, a norm coming from an inner product satisfies the *parallelogram identity*: $\|x+y\|^2+\|x-y\|^2 = 2(\|x\|^2+\|y\|^2)$; it can be shown (see [19, p. 130]) that only a norm which satisfies the parallelogram identity comes from an inner product. We now use the inner product to define the concept of orthogonal vectors.

Definition 2.5. Let X be a vector space with inner product $\langle \cdot, \cdot \rangle$. We say that vectors $x, y \in X$ are orthogonal if

$$\langle x, y \rangle = 0.$$

An important result about inner product spaces is the *Cauchy-Schwarz* inequality.

Proposition 2.1 (Cauchy-Schwarz). *In an inner product space X, the following inequality holds for all $x, y \in X$:*

$$| \langle x, y \rangle | \leq \|x\| \|y\|$$

with equality if and only if x and y are linearly dependent.

Here are some examples of normed and inner product spaces.

Example 2.1. *Consider the vector space $C[a, b]$ of all complex-valued continuous functions on the interval $[a, b]$. For any real number $p \geq 1$ we can define a norm on $C[a, b]$ by*

$$\|x\|_p := \left(\int_a^b |x(t)|^p \right)^{1/p}. \tag{2.2}$$

The following also defines a norm on $C[a, b]$:

$$\|x\|_\infty := \max_{t \in [a,b]} |x(t)|. \tag{2.3}$$

So $(C[a, b], \| \cdot \|_p)$ as well as $(C[a, b], \| \cdot \|_\infty)$ are normed spaces. Only one of these is an inner product space however since only one of the above norms comes from an inner product, namely the 2-norm. It is the only one which satisfies the parallelogram identity. Its inner product is given by

$$\langle x, y \rangle := \int_a^b x(t)\overline{y(t)}dt. \tag{2.4}$$

If we restrict our space to real-valued functions, the definition of the inner product remains the same except that there is no need to conjugate the second argument inside the integral.

2.2 Metric Spaces (Review)

Definition 2.6. A *metric space* is a set X together with a *distance function* (or a *metric*) $d : X \times X \to \mathbb{R}$ satisfying the following properties for all $x, y, z \in X$:

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

In a normed vector space X, we can use the norm on X to define a distance function by $d(x, y) := \|x - y\|$. This metric is said to be *induced* by the norm on X. Hence, all normed spaces are also metric spaces, and in fact, all inner product spaces are metric spaces because the inner product can also be used to induce a metric, by inducing a norm first.

Definition 2.7. A sequence (x_n) in a metric space X is called *a Cauchy sequence* if

$$\lim_{m,n\to\infty} d(x_m, x_n) = 0.$$

Definition 2.8. A metric space X in which every Cauchy sequence converges is said to be *complete*.

2.3 Banach and Hilbert Spaces

Definition 2.9. A *Banach space* is a normed vector space X which is complete with respect to the metric induced by its norm.

Definition 2.10. A *Hilbert space* is an inner product space which is complete with respect to the metric induced by its inner product.

Hence a Hilbert space is also a Banach space. We are mainly interested in Hilbert spaces since the *CG* algorithm requires the use of an inner product. As mentioned earlier, Hilbert spaces resemble greatly the standard vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ because of their similar vector-space and metric-space structure. The main difference is that a general Hilbert space may be infinite dimensional. Here are some examples of Banach and Hilbert spaces other than $\mathbb{R}^n$ and $\mathbb{C}^n$.

Example 2.2. *Consider the complex space $(C[a,b], \|\cdot\|_p)$ from Example 2.1. It is not complete; we can complete it, however, by constructing a limit function corresponding to every Cauchy sequence of continuous functions on $[a,b]$. Thus we can obtain the complete "version" of the space denoted by $L^p[a,b]$. Hence $L^p[a,b]$ is a Banach space. In the language of measure theory, it contains all Lebesgue-measurable functions which are p^{th}-power integrable on $[a,b]$. More formally, $L^p[a,b]$ contains equivalence classes of Lebesgue-measurable functions which are p^{th}-power integrable on $[a,b]$, where two functions are from the same class provided they are equal almost everywhere. In other words, two functions are from the same class, if they only differ on a set of Lebesgue-measure zero. Since only the $2-$norm comes from an inner product, the only Hilbert space obtained by the above procedure is $L^2[a,b]$.*

Example 2.3. *Consider the space $l^2(\mathbb{N})$, of all complex sequences (x_n) which are square summable, with inner product defined by*

$$\langle x, y\rangle := \sum_{i=1}^{\infty} x_i \overline{y_i} \tag{2.5}$$

where $x = (x_1, x_2, \cdots)$ and $y = (y_1, y_2, \cdots)$. The norm of x in this space is just

$$\|x\| := \left(\sum_{i=1}^{\infty} |x_i|^2\right)^{1/2} \tag{2.6}$$

This summation converges because x is square summable. Moreover, the summation in the definition of the inner product also converges by the Cauchy-Schwarz inequality. The space $l^2(\mathbb{N})$ is in fact complete ([15, p. 20]) and hence, a Hilbert space.

Equipped with some basic knowledge and examples of Hilbert spaces, we are now ready to discuss linear operators defined on these.

2.4 Linear Operators on Hilbert Spaces

Definition 2.11. Let H be a Hilbert space over $\mathbb{C}$. A *linear operator* on H is a mapping $L : H \to H$ such that

$$L(ax + by) = aL(x) + bL(y) \tag{2.7}$$

for all $x, y \in H$ and scalars a, b.

We usually omit the parentheses for single arguments and write Lx instead of $L(x)$. Sometimes, the domain of a linear operator is not the whole space H but some subspace $\mathcal{D} \subset H$. In this case we write $L : \mathcal{D} \to H$.

Definition 2.12. Let $L : \mathcal{D} \to H$ be a linear operator where $\mathcal{D}$ is a subspace of a Hilbert space H. We say that L is *bounded* on $\mathcal{D}$ if there is a $c \in \mathbb{R}$ such that

$$\|Lx\| \leq c\|x\| \tag{2.8}$$

for all $x \in \mathcal{D}$. In that case we define the *norm* of L to be the smallest c for which the above inequality holds, that is,

$$\|L\| = \sup_{x \in \mathcal{D}, x \neq 0} \frac{\|Lx\|}{\|x\|}. \tag{2.9}$$

If L is bounded, an equivalent way to define its norm is by

$$\|L\| = \sup_{x \in \mathcal{D}, \|x\|=1} \|Lx\|. \tag{2.10}$$

It is interesting to note that a bounded operator can be easily extended from a dense subspace of H to the whole space without changing the norm of the operator.

This is important because often we are interested in the properties of a linear operator on some inner product space which is not complete, say, the spaces defined in Example 2.1. However, to avoid certain difficulties, such as non-convergent Cauchy sequences, and to be able to use all the Hilbert space theory already in place, we need to complete our space. The following proposition asserts that all bounded operators can be easily extended to the completion of the original space without changing their norm. In particular, the extended operator will still be bounded.

Proposition 2.2 (Bounded extension). *Let $L : \mathcal{D} \to H$ be a bounded linear operator where $\mathcal{D}$ is a subspace of a Hilbert space H. Then L has an extension $\tilde{L} : \overline{\mathcal{D}} :\to H$ such that $\tilde{L}$ is also bounded and $\|\tilde{L}\| = \|L\|$.*

Proof. [19, p.100] □

Here are some examples of bounded and unbounded linear operators.

Example 2.4. *The identity operator $I : H \to H$ and the zero operator $0 : H \to H$ are bounded on any Hilbert space H. The first one satisfies $\|Ix\| = \|x\|$ for all $x \in H$ and hence has a norm 1. The second satisfies $\|0x\| = \|0\|$ and hence has norm 0.*

Example 2.5. *The right-shift operator R on $l^2(\mathbb{N})$ is bounded because*

$$\|Rx\|^2 = 0^2 + x_1^2 + x_2^2 + \cdots = x_1^2 + x_2^2 + \cdots = \|x\|^2$$

which implies that $\|Rx\| = \|x\|$ for all $x \in l^2(\mathbb{N})$. Hence the operator has norm 1.

Example 2.6. *The integral operator on $L^2[0,1]$ defined by*

$$Lx(t) = \int_0^t x(s)ds \tag{2.11}$$

is bounded because for any $x \in L^2[0,1]$ with$\|x\| = 1$, we have

$$\|Lx\|^2 = \int_0^1 \left(\int_0^1 |k(s)x(s)|ds \right)^2 dt,$$

where $k(s)$ is the characteristic function of $[0,t]$. By the Cauchy-Schwarz inequality, Proposition 2.1, the expression on the right is less than or equal to

$$\int_0^1 \left(\int_0^1 |k(s)|^2 ds \int_0^1 |x(s)|^2 ds \right) dt \leq \|x\|^2 \int_0^1 \int_0^1 |k(s)|^2 ds\, dt = \frac{1}{2}\|x\|^2.$$

Example 2.7. *The differential operator defined by*

$$Lx(t) = x'(t) \tag{2.12}$$

on the subspace of $L^2[0,1]$ consisting of differentiable functions is unbounded. Indeed, let $x_n(t) = (2n+1)\, t^n$, where $n \in \mathbb{N}$. Then $\|x_n\| = 1$ for all n, by integration. On the other hand, $Lx_n(t) = (2n+1)n\, t^{n-1}$ with $\|Lx_n\| = \sqrt{\frac{(2n+1)n}{2n-1}}$ which is unbounded for an arbitrary n. A similar proof is found in the book by Kreyszig [19, p.92], but for the ∞-norm.

Example 2.8. *Any linear operator A on $\mathbb{C}^n$ is bounded. To see this, let $e_1, e_2, \ldots e_n$ be the standard basis for $\mathbb{C}^n$, and let $x = \sum_{j=1}^n \alpha_j e_j$ have norm 1. In particular, each α_j has magnitude less than 1. Applying A to x and taking the norm yields*

$$\|Ax\| = \sum_{j=1}^n |\alpha_j| \|Ae_j\| \leq \max_{j\in\{1\ldots n\}} \|Ae_j\| \sum_{j=1}^n |\alpha_j| \leq n \max_{j\in\{1\ldots n\}} \|Ae_j\|.$$

Hence,

$$\|A\| \leq n \max_{j\in\{1\ldots n\}} \|Ae_j\|.$$

The same proof can be used to show that a linear operator on any finite dimensional Hilbert space is bounded.

We now extend the basic notion of *adjoint* and *self-adjoint* matrices to linear operators on a Hilbert space.

Definition 2.13 (Adjoint). Let $L : H \to H$ be a bounded linear operator on a Hilbert space H. We define the *adjoint* operator of L as the operator $L^* : H \to H$ satisfying

$$\langle Lx, y\rangle = \langle x, L^* y\rangle \tag{2.13}$$

for all $x, y \in H$.

From the definition, we see that adjoint operators in Hilbert spaces play a similar role as adjoint matrices in $\mathbb{C}^n$. Note that the above definition already assumes the existence of the adjoint operator. The following proposition asserts that this definition indeed makes sense.

Proposition 2.3. *The adjoint operator L^* in Definition 2.13 exists, is unique, and is a bounded linear operator with norm*

$$\|L^*\| = \|L\|.$$

Proof. [19, p.196]. □

We now list some useful properties of adjoint operators. As the reader will no doubt notice these are exactly the same as for matrices. The proofs are the same as for the matrix case and can be found in any basic linear algebra book. For example see [20, p. 403].

Proposition 2.4 (Properties of adjoint operators). *Let L_1 and L_2 be bounded linear operators on a Hilbert space H. Then*

1. $(L_1^*)^* = L_1$,
2. $(L_1 + L_2)^* = L_1^* + L_2^*$,
3. $(cL_1)^* = \overline{c}L_1^*$,
4. $(L_1 L_2)^* = L_2^* L_1^*$.

Definition 2.14 (Self-adjoint operator). Let L be a bounded, linear operator on a Hilbert space H. Then L is said to be *self-adjoint* if $L^* = L$, that is, if

$$\langle Lx, y\rangle = \langle x, Ly\rangle$$

for all $x, y \in H$.

Proposition 2.5. *Let L_1 and L_2 be bounded self-adjoint operators on a Hilbert space H. Then L_1+L_2 is self-adjoint, and L_1L_2 is self adjoint provided L_1 and L_2 commute.*

Proof. The result is an immediate consequence of the definition of a self-adjoint operator (2.14) and statements 2 and 4 of Proposition 2.4. □

For self-adjoint linear operators on a Hilbert space H, we have the following remarkable property.

Proposition 2.6. *Let L be a self-adjoint linear operator on a Hilbert space H. Then $\langle Lx, x\rangle$ is real for all $x \in H$.*

Proof. By conjugate symmetry (Definition 2.1) and the definition of a self-adjoint operator (Definition 2.14), we have $\overline{\langle Lx, x\rangle} = \langle x, Lx\rangle = \langle Lx, x\rangle$, which proves the proposition. □

Definition 2.15. Let L be a bounded self-adjoint linear operator on a Hilbert space H. Then L is called *positive* if $\langle Lx, x\rangle > 0$ for all $x \in H$, $x \neq 0$.

Proposition 2.7. *Let L_1 and L_2 be bounded, self-adjoint positive linear operators on a Hilbert space H. Then $L_1 + L_2$ is positive and L_1L_2 is positive provided L_1 and L_2 commute.*

Proof. From Proposition 2.5, $L_1 + L_2$ is self-adjoint. Also, for $x \in H$ positivity of L_1 and L_2 implies

$$\langle (L_1 + L_2)\, x, x\rangle = \langle L_1 x, x\rangle + \langle L_2 x, x\rangle \geq 0$$

The proof of the second part of the proposition is rather complicated and can be found in [19, pp. 470–473]. □

Proposition 2.8 (Uniform Boundedness Theorem). *Let A_n be a sequence of bounded linear operators on a Hilbert space H such that $\|A_n x\|$ is bounded for every fixed $x \in H$, say $\|A_n x\| \leq c_x$ for $n = 0, 1, \ldots$. Then A_n is uniformly bounded, that is, there exists a constant $c \in \mathbb{R}$ such that $\|A_n\| \leq c$ for $n = 0, 1, \ldots$.*

Proof. See [19, pp. 249–250]. □

2.5 Spectral Theory of Linear Operators

Spectral theory of linear operators is essentially an extension of matrix eigenvalue theory. Complications arise, however, due to the fact that the underlying Hilbert space may be infinite-dimensional. This leads to the need of defining spectral values other than eigenvalues. We mention that the theory described here is valid even if one replaces the Hilbert space H by a general normed space X.

Definition 2.16 (inverse). Let $L : \mathcal{D} \to H$ be a linear operator. We say that L^{-1} exists if L is injective, that is, if $Lx = Ly$ implies $x = y$.

Proposition 2.9. *Let $L : \mathcal{D} \to H$ be a linear operator. If L^{-1} exists, then it is a linear operator with $L^{-1} : L(\mathcal{D}) \to \mathcal{D}$.*

Proof. See [19, pp. 87–88]. □

Definition 2.17 (Resolvent). Let $L : \mathcal{D} \to H$ be a linear operator where $\mathcal{D}$ is a subspace of a Hilbert space H, and let $\lambda \in \mathbb{C}$. With L we associate the operator $L_\lambda := (L - \lambda I)$, where I is the identity operator. If L_λ^{-1} exists we call it the *resolvent* of L corresponding to λ.

Note the similarity between the above construction of L_λ and the operator used to construct the characteristic polynomial in the matrix case. We now use the definition of resolvent to define the concept of spectrum of a linear operator (see [19, p.371]). Alternative definitions of spectrum are given in [18, p. 374] and [15, p. 50]. The reason for the different definitions of spectrum in the literature is that L^{-1} is not always defined as in (Definition 2.16). Nevertheless, all definitions of spectrum are equivalent.

Definition 2.18 (Spectrum). Let $L : \mathcal{D} \to H$ be a linear operator where $\mathcal{D}$ is a subspace of a Hilbert space H. Then:

1. The *point spectrum* of L is the set

$$\sigma_p(L) = \{\lambda \in \mathbb{C} : L_\lambda^{-1} \textit{ does not exist}\}. \tag{2.14}$$

2. The *continuous spectrum* of L is the set

$$\sigma_c(L) = \{\lambda \in \mathbb{C} : L_\lambda^{-1} \textit{ exists, has a dense domain in } H, \textit{ but is unbounded}\}. \tag{2.15}$$

3. The *residual spectrum* of L is the set

$$\sigma_r(L) = \{\lambda \in \mathbb{C} : L_\lambda^{-1} \textit{ exists, but its domain is not dense in } H\}. \tag{2.16}$$

The union of these three sets is called the *spectrum* of L and is denoted by $\sigma(L)$.

Definition 2.19 (Spectral value, eigenvalue, regular value). Let $L : \mathcal{D} \to H$ be a linear operator where $\mathcal{D}$ is a subspace of a Hilbert space H. If a complex number $\lambda \in \sigma(L)$ then λ is called a *spectral value* of L. In particular, if $\lambda \in \sigma_p(L)$ then it is called an eigenvalue of L. If $\lambda \notin \sigma(L)$, then it is called a *regular value* of L.

Notice that the above definition of eigenvalue is equivalent to the standard definition since

$$Lx = \lambda x$$

for some $x \neq 0$, if and only if

$$(L - \lambda I)\, x = 0,$$

for some $x \neq 0$, if and only if $(L - \lambda I)$ is not injective, and hence not invertible. From the definition of spectrum, we see that the spectral values of an operator are not necessarily eigenvalues. One reason for the more complicated definition of spectrum in

Hilbert spaces is that even though $(L-\lambda I)$ may be invertible, the inverse may possess some "pathological" properties, such as the ones described in 2 and 3 of Definition 2.18, which need special attention. That is why the numbers $\lambda \in \mathbb{C}$ for which the inverse exists but has these "bad" properties cannot be considered as regular values and have to be included in the spectrum, even though they are not eigenvalues. In $\mathbb{R}^n$ or $\mathbb{C}^n$ this problem does not occur since these spaces are finite dimensional. Hence, if L_λ is invertible then it is injective, and hence bijective, allowing L_λ^{-1} to have the whole space as domain. Moreover L_λ^{-1} will be a bounded operator since any linear operator on a finite dimensional space is bounded by Example 2.8. Here are some examples of operators with spectral values other than eigenvalues.

Example 2.9. *Let $H = L^2[1,2]$ and let L be defined by $(Lx)(t) = tx(t)$. Then L has no eigenvalues because there is no $x(t) \in H$ such that $(Lx)(t) = tx(t)$. However, for any $\lambda \in [1,2]$, the resolvent L_λ is unbounded implying that these λ are spectral values of L.*

The following example is taken from [19, p. 372].

Example 2.10. *Let $H = l^2(\mathbb{N})$ and let R denote the right-shift operator. Then R_0 is invertible, the inverse being the left-shift operator $R_0^{-1} : E \to H$, where $E = \{x \in H : x_1 = 0\}$. Although R_0^{-1} exists and is bounded (its norm is in fact 1), its domain E is not dense in H and hence 0 belongs to the residual spectrum σ_r of R.*

We end this chapter with the following simple but useful observation which we state as a proposition.

Proposition 2.10. *Let L be a linear operator on a Hilbert space H. Then $\sigma(aL + bI) = a\sigma(L) + b$.*

Proof. The proof is an immediate consequence of the definition of the spectrum and the linearity of the inner product in the first component (Definition 2.1). □

Chapter 3

Conjugate Gradient in Hilbert Spaces

In this chapter, we extend the application of the *CG* algorithm to general Hilbert spaces. We first show that a self-adjoint positive operator can be used to define an inner product on H other than the original one. We then examine some important results concerning the spectral properties of such operators. Finally, we define the concept of a *self-adjoint positive definite* linear operator and we use it to prove a convergence result for *CG* similar to the one already given in Chapter 1 (Theorem 1.8).

Proposition 3.1 (L-inner product). *Let L be a bounded self-adjoint positive linear operator on a Hilbert space H. Then the mapping $\langle \cdot, \cdot \rangle_L : H \times H \to \mathbb{C}$ defined by*

$$\langle x, y \rangle_L := \langle Lx, y \rangle \tag{3.1}$$

defines an inner product on H.

Proof. For all $x, y, z \in H$ and $a, b \in \mathbb{C}$ we have:

1. **linearity:**

Using linearity of the original inner product and of the operator L we have

$$\begin{aligned}\langle ax+by,z\rangle_L &= \langle L(ax+by),z\rangle \\ &= \langle aLx+bLy,z\rangle \\ &= a\langle Lx,z\rangle + b\langle Ly,z\rangle \\ &= a\langle x,z\rangle_L + b\langle y,z\rangle_L .\end{aligned}$$

2. **conjugate symmetry:**

   Using conjugate symmetry of the original inner-product and the fact that L is self-adjoint, we have

$$\langle x,y\rangle_L = \langle Lx,y\rangle = \overline{\langle y,Lx\rangle} = \overline{\langle Ly,x\rangle} = \overline{\langle y,x\rangle}_L.$$

3. **positive-definiteness:**

   This property of the $L-$inner product follows directly from the definition of positivity (Definition 2.15).

□

3.1 Spectral Properties of Self-Adjoint Positive Linear Operators

Theorem 3.1 (Spectrum). *The spectrum $\sigma(L)$ of a bounded self-adjoint linear operator L on a Hilbert space H is real.*

Proof. [19, pp. 463-464]. □

Theorem 3.2 (Spectrum). *The spectrum $\sigma(L)$ of a bounded self-adjoint linear operator L on a Hilbert space H lies in the closed interval $[m,M]$ where*

$$m := \inf_{x\in H, x\neq 0} \frac{\langle Lx,x\rangle}{\langle x,x\rangle} = \inf_{\|x\|=1} \langle Lx,x\rangle , \tag{3.2}$$

$$M := \sup_{x \in H, x \neq 0} \frac{\langle Lx, x \rangle}{\langle x, x \rangle} = \sup_{\|x\|=1} \langle Lx, x \rangle . \tag{3.3}$$

Moreover, m and M are spectral values of L.

Proof. [19, pp. 465-467]. □

This theorem motivates the following definition of a self-adjoint positive-definite operator.

Definition 3.3. Let L be a bounded self-adjoint positive linear operator on a Hilbert space H. Let m be as in the previous theorem. If $m > 0$ we say that L is *positive definite.*

Hence, we distinguish the concepts of a positive operator (Def 2.15) and a positive definite operator (Def 3.3). A positive definite operator clearly satisfies Definition 2.15 and hence is positive. The converse is false in general since a positive operator may have 0 as a spectral value, though not as an eigenvalue. This problem does not occur in $\mathbb{C}^n$ since the spectral values there are precisely the eigenvalues. In that space, Definitions 2.15 and 3.3 are equivalent. Since our objective, in this thesis, is to adapt the proof of convergence of CG in $\mathbb{C}^n$ (Theorem 1.8) to an arbitrary Hilbert space, we need operators whose spectrum is away from zero, namely positive definite operators.

An important property of a self-adjoint positive definite linear operator $L : H \to H$ is that the equation $Lu = f$ always has a unique solution. Since $0 \notin \sigma(L)$, L is invertible, bounded, and has a dense set as domain; thus, the whole H by the bounded extension theorem (Proposition 2.2) making L a bijection. Hence, when applying CG to this kind of operators, we do not need to worry about questions of existence and uniqueness of the solution. The solution exists and is unique!

Theorem 3.4 (Norm and Spectrum). *Let L be a bounded self-adjoint linear operator on a Hilbert space H. Then*

$$\|L\| = \max(|m|, |M|) \tag{3.4}$$

where m and M are as in the previous theorem.

Proof. See [19, p. 466]. □

Hence, if L is positive, then we simply have $\|L\| = M$.

Definition 3.5. Let L_1 and L_2 be bounded self-adjoint operators on a Hilbert space H. We say that $L_1 \leq L_2$ if $\langle L_1 x, x\rangle \leq \langle L_2 x, x\rangle$ for all $x \in H$.

A strict inequality is defined similarly. Note that $L_1 \leq L_2$ if and only if $0 \leq L_2 - L_1$. The following three results are fundamental in our proof of convergence of CG whereas the fourth is only used in the last section of our thesis.

Theorem 3.6. *Let A_n be a sequence of bounded self-adjoint linear operators on a complex Hilbert space H such that*

$$A_1 \leq A_2 \leq \ldots \leq A_n \ldots \leq K \tag{3.5}$$

where K is a bounded self-adjoint linear operator on H. Suppose that any A_j commutes with K and with every A_m. Then, there exists a bounded self-adjoint linear operator A such that $A_n x \to Ax$ for all $x \in H$ satisfying $A \leq K$.

Proof. Also see [19, pp.473–474]. Consider $S_n = K - A_n$. We first show that the sequence $(\langle S_n^2 x, x\rangle)$ converges for every $x \in H$. S_n is self-adjoint since A_n and K are self-adjoint. We have,

$$S_m^2 - S_n S_m = (S_m - S_n) S_m = (A_n - A_m)(K - A_m).$$

Let $m < n$. Then, $A_n - A_m$ and $K - T_m$ are positive by Definition 3.5. Since these operators commute, their product is positive by Proposition 2.7. Hence, $S_m^2 \geq S_n S_m$. Similarly, $S_n S_m \geq S_n^2$. Putting these together yields

$$S_m^2 \geq S_n S_m \geq S_n^2 \tag{3.6}$$

implying that $\langle S_n^2 x, x\rangle$ is a decreasing sequence of real numbers, by self-adjointness of S_n, for any fixed $x \in H$. Moreover this sequence is bounded below by 0 since $\langle S_n^2 x, x\rangle = \langle S_n x, S_n x\rangle \geq 0$. Therefore $\langle S_n^2 x, x\rangle$ converges.

We now show that $A_n x$ converges for a fixed $x \in H$. By assumption, every A_n commutes with every A_m as well as with K. Hence the S_n's also commute. Now, Equation 3.6 implies that, for $m < n$, $-2\langle S_n S_m x, x\rangle \leq -2\langle S_n^2 x, x\rangle$. Thus,

$$\begin{aligned}
\|S_m x - S_n x\|^2 &= \langle (S_m - S_n)\, x, (S_m - S_n)\, x\rangle \\
&= \left\langle (S_m - S_n)^2\, x, x\right\rangle \\
&= \left\langle S_m^2 x, x\right\rangle - 2\langle S_n S_m x, x\rangle + \left\langle S_n^2 x, x\right\rangle \\
&\leq \left\langle S_m^2 x, x\right\rangle - \left\langle S_n^2 x, x\right\rangle .
\end{aligned}$$

From this and the convergence of $\langle S_n^2 x, x\rangle$, we deduce that $S_n x$ is Cauchy. By completeness of H this sequence converges. Hence $A_n x = (K - S_n)x$ also converges and the mapping

$$x \mapsto \lim_{n\to\infty} A_n x$$

defines a linear operator $A : H \to H$. This operator is bounded by the uniform boundedness theorem (Proposition 2.8) because $A_n x$ converges so that it is bounded for every fixed $x \in H$ and $n = 1, 2, \ldots$. Also, A is self-adjoint because the A_n are self-adjoint and the inner product is continuous. Finally, $A \leq K$ follows from $A_n \leq K$.

□

Theorem 3.7 (Positive square root). *Let L be a bounded self-adjoint positive linear operator on a Hilbert space H. Then L has a unique self-adjoint positive square root, that is, there is a unique bounded self-adjoint positive linear operator A on H such that $A^2 = L$. Moreover, A commutes with every bounded linear operator on H which commutes with L.*

Proof. Our proof follows closely the one given in [19, pp. 476–479]. We proceed in three steps. First we show that the square root exists under the additional assumption $L \leq I$ where I denotes the identity operator. We then prove uniqueness of the square root. Finally we show that the result holds when the assumption in step one is removed.

Assume for the moment that $L \leq I$. We construct a sequence of operators A_n converging to an operator A with the required properties. Consider $A_0 := 0$ and

$$A_{n+1} = A_n + \frac{1}{2}\left(L - A_n^2\right) \text{ for } n = 0, 1, \ldots. \tag{3.7}$$

Each A_n is a real polynomial in L. Hence the A_n's are self-adjoint and commute with every operator that L commutes with. We now show that

$$A_n \leq I. \tag{3.8}$$

We have $A_0 \leq I$. Let $n > 0$. Since $I - A_n$ is self-adjoint, then $(I - A_{n-1})^2 \geq 0$. Also, $I - L \geq 0$ by assumption. Thus,

$$\begin{aligned} 0 &\leq \frac{1}{2}(I - A_{n-1})^2 + \frac{1}{2}(I - L) \\ &= I - A_{n-1} - \frac{1}{2}\left(L - A_{n-1}^2\right) \\ &= I - A_n. \end{aligned}$$

Next, we use (3.8) and induction to show that

$$0 < A_n \leq A_{n+1}, \; n = 1, 2, \ldots. \tag{3.9}$$

Equation (3.7) gives $A_0 = 0 \leq \frac{1}{2}L = A_1$. Assume $A_{n-1} \leq A_n$ for some n. Then using the fact that A_n and A_{n-1} commute, we have

$$\begin{aligned} A_{n+1} - A_n &= A_n + \frac{1}{2}\left(L - A_n^2\right) - A_{n-1} - \frac{1}{2}\left(L - A_{n-1}^2\right) \\ &= A_n - A_{n-1} - \frac{1}{2}\left(A_n^2 - A_{n-1}^2\right) \\ &= (A_n - A_{n-1})\left[I - \frac{1}{2}(A_n + A_{n-1})\right]. \end{aligned}$$

Here, $(A_n - A_{n-1}) \geq 0$ by the induction hypothesis and $[I - \frac{1}{2}(A_n + A_{n-1})] \geq 0$ by (3.8). Hence $A_{n+1} - A_n \geq 0$ by Proposition 2.7. Note also that for $n > 0$ we have $0 < \frac{1}{2}L = A_1 \leq A_n$.

Now, Equation (3.9) and the fact that A_n commutes with A_m as well as with I imply the existence of a bounded self-adjoint linear operator A such that $A_n x \to Ax$ for all $x \in H$ by Theorem 3.6. In particular, since $A_n x$ converges, (3.7) gives

$$A_{n+1}x - A_n x = \frac{1}{2}\left(Lx - A_n^2 x\right) \to 0$$

as $n \to \infty$. Hence, $Lx - A^2x = 0$ for all $x \in H$, that is, $L = A^2$. Also, since for $n > 1$, $0 < A_n$ then $\langle A_n x, x\rangle > 0$ implying $\langle Ax, x\rangle > 0$ for every $x \in H$, $x \neq 0$ by the continuity of the inner product. Hence A is positive.

We now show that A commutes with every bounded operator S on H which commutes with L. From an earlier remark, since A_n is a polynomial in L then A_n commutes with S for each $n = 0, 1, \ldots$. Thus, for every $x \in H$ we have,

$$ASx = \lim_{n\to\infty} A_n(Sx) = \lim_{n\to\infty} S(A_n x) = S \lim_{n\to\infty} A_n x = SAx.$$

The operator S can be put outside the limit sign because it is bounded, hence continuous.

Next, we show that A is unique. Suppose that A is the square root of L constructed above and B is another positive square root of L. Then, $BL = BB^2 = B^2B = LB$ implying that B commutes with L. Hence B also commutes with A. Let $x \in H$ and $y = (A - B)x$. Then $\langle Ay, y\rangle \geq 0$ and $\langle By, y\rangle \geq 0$, by positivity of A and B, with equality if and only if $y = 0$. Hence,

$$\langle Ay, y\rangle + \langle By, y\rangle = \langle (A + B)y, y\rangle = \left\langle (A^2 - B^2)x, y\right\rangle = 0$$

which can only occur if $\langle Ay, y\rangle = \langle By, y\rangle = 0$ because both terms on the left are positive. Hence, $y = 0$ implying that $Ax = Bx$ for all $x \in H$ yielding $A = B$. We point out that commutativity of A and B was needed in order to for the factorization $A^2 - B^2 = (A - B)(A + B)$ to hold.

Finally, the theorem also holds without the original assumption $L \leq I$ because L is bounded. We then have that the unique positive square root of L is just

$$L^{\frac{1}{2}} = \|L\|^{\frac{1}{2}} \left(\frac{L}{\|L\|}\right)^{\frac{1}{2}},$$

where $\frac{L}{\|L\|} \leq I$. □

Theorem 3.8. *Let L be a bounded self-adjoint positive operator on a Hilbert space H and let p be a polynomial with real coefficients. Then*

$$\|p(L)\| \leq \max_{\lambda \in [m,M]} |p(\lambda)| \tag{3.10}$$

Proof. The proof can be found in [19, p.509]. We give it here for completeness. First, we show that if $q(\lambda) \geq 0$ for all $\lambda \in [m, M]$, for a real polynomial q, then $q(L) \geq 0$. Note that $q(L)$ is self-adjoint because L is self-adjoint and q is real. The idea is to factor q so that $q(L)$ is realized as a product of commuting positive operators. The details are as follows. Since q has real coefficients, complex zeros, if any, must occur in conjugate pairs. Since q changes sign if λ passes through a zero of odd multiplicity,

and $q(\lambda) \geq 0$ on $[m, M]$, then zeros of q in (m, M) must be of even multiplicity. Hence we can write

$$q(\lambda) = \alpha \prod_j (\lambda - \beta_j) \prod_k (\gamma_k - \lambda) \prod_l \left[(\lambda - \mu_l)^2 + \nu_l^2\right] \tag{3.11}$$

where $\beta_j \leq m$ and $\gamma_k \geq M$ and the quadratic factors correspond to complex conjugate zeros and to real zeros in (m, M), the latter occuring when $\nu_l = 0$. Written in this form, we see that each term in q is positive for $\lambda \in [m, M]$. Since $q(\lambda) \geq 0$ on that interval, by assumption, then $\alpha > 0$ for $q \neq 0$. If $q = 0$, the result trivially holds. We now replace λ by L in Equation (3.11). Then each of the factors is a self-adjoint positive operator by Definition 3.5 and Theorem 3.2. In particular, each $L - \mu_l I$ is self-adjoint so that its square is positive implying that

$$(L - \mu_l I)^2 + \nu_l^2 I \geq 0.$$

Moreover, all operators in the product commute. Therefore, $q(L) \geq 0$ by Proposition 2.7.

We are now ready to prove the statement of the theorem. Let k denote the maximum of $|p(\lambda)|$ on $[m, M]$. Then $k^2 - p(\lambda)^2 \geq 0$ on $[m, M]$. Hence the above result yields $k^2 I - p(L)^2 \geq 0$, that is, $p(L)^2 \leq k^2 I$ implying that for all $x \in H$,

$$\|P(L)x\|^2 = \langle p(L)x, p(L)x \rangle = \left\langle p(L)^2 x, x \right\rangle \leq k^2 \langle x, x \rangle .$$

Equation (3.10) now follows by taking square roots and then the supremum over all $x \in H$. □

Theorem 3.9. *Let L be a bounded self-adjoint positive definite linear operator on a Hilbert space H, with $\sigma(L) \subseteq [m, M]$ where $m > 0$. Then L^{-1} is self-adjoint positive definite with $\sigma(L^{-1}) \subseteq [\frac{1}{M}, \frac{1}{m}]$.*

Proof. Take a deep breath. Here it goes .... From an earlier discussion (p. 35) we know that L is a bijection from $H \to H$ and L^{-1} is a bounded operator. Let $Lx = \bar{x}$ and $Ly = \bar{y}$. This implies,

$$\langle Lx, y\rangle = \left\langle \bar{x}, L^{-1}\bar{y}\right\rangle .$$

On the other hand, since L is self-adjoint, we have

$$\langle Lx, y\rangle = \langle x, Ly\rangle = \left\langle L^{-1}\bar{x}, \bar{y}\right\rangle ,$$

showing that L^{-1} is self-adjoint. Moreover, it is positive, since

$$\left\langle L^{-1}\bar{x}, \bar{x}\right\rangle = \langle x, Lx\rangle > 0$$

by the positivity of L. In fact, $0 \notin \sigma(L^{-1})$ by the definition of the spectrum (Definition 2.18) since $(L^{-1})^{-1} = L$ exists, is bounded and has H as domain. Thus L^{-1} is positive definite (Theorem 3.2). To show the last part of the result, note that L^{-1} has a self-adjoint positive definite square root from Theorem 3.7. Therefore, for $x \in H$, we have, by Theorem 3.2,

$$\frac{1}{m} \geq \frac{\langle x, x\rangle}{\langle Lx, x\rangle} = \frac{\langle L^{-1}\bar{x}, L^{-1}\bar{x}\rangle}{\langle \bar{x}, L^{-1}\bar{x}\rangle} = \frac{\left\langle L^{-1}\left(L^{-\frac{1}{2}}\bar{x}\right), \left(L^{-\frac{1}{2}}\bar{x}\right)\right\rangle}{\left\langle \left(L^{-\frac{1}{2}}\bar{x}\right), \left(L^{-\frac{1}{2}}\bar{x}\right)\right\rangle}.$$

Since $L^{-\frac{1}{2}}$ is self-adjoint positive definite, then it is a bijection from $H \to H$; hence, $\left(L^{-\frac{1}{2}}\bar{x}\right)$ can take any value in H. Theorem 3.2 now implies that $\frac{1}{m}$ is an upper bound of $\sigma(L^{-1})$. A similar argument shows that $\frac{1}{M}$ is a lower bound of $\sigma(L^{-1})$. □

3.2 Convergence of CG in Hilbert Spaces

We now prove convergence of *CG* for bounded self-adjoint positive definite linear operators on a Hilbert space H. The proof is by rate of convergence. We advise the reader to go back and review the main steps of the proof given in Chapter 1 for the case

$\mathbb{C}^n$. We recall that P_j denotes the set of real polynomials p of degree at most j such that $p(0) = 1$ (Definition 1.7). We also recall that by Lemma 1.1 the error of the j^{th} iteration of CG can be written as $e_j = p_j(L)e_0$ for some $p_j \in P_j$. The proof of this fact on a general Hilbert space is identical to that of Lemma 1.1. Hence on an arbitrary Hilbert space, CG finds a polynomial $p_j \in P_j$ such that $\|p_j(L)e_0\|_L$ is minimal. We now have the following equivalent of Lemma 1.2. In the finite dimensional case, the proof of the corresponding result uses an orthonormal basis for $\mathbb{C}^n$ formed using the eigenvectors of the matrix A. Here we cannot use such a result since the underlying Hilbert space may be infinite. We prove the result using our knowledge of spectral theory and square roots.

Lemma 3.1. *Consider the problem $Lu = f$ where L is a bounded self-adjoint positive definite linear operator on a Hilbert space H. If at step j the* CG *iteration applied to this problem has not already converged, then*

$$\|e_j\|_L = \inf_{p \in P_j} \|p(L)e_0\|_L \leq \left(\inf_{p \in P_j} \max_{\lambda \in \sigma(L)} |p(\lambda)| \right) \|e_0\|_L \tag{3.12}$$

Proof. The equality is just a restatement of Lemma 1.1 and holds for an arbitrary Hilbert space. We now prove the inequality. Since L is self-adjoint positive-definite, its spectrum lies in $[m, M]$ where $m > 0$. By Theorem 3.7, L has a unique self-adjoint positive square root $L^{1/2}$. Now, consider an arbitrary polynomial $p \in P_j$. Note that the operators $L^{1/2}$ and $p(L)$ commute because the latter could be rewritten as a polynomial in $L^{1/2}$. We have

$$\begin{aligned}
\|p(L)e_0\|_L^2 &= \langle Lp(L)e_0, p(L)e_0\rangle \\
&= \left\langle L^{1/2}p(L)e_0, L^{1/2}p(L)e_0\right\rangle \\
&= \|L^{1/2}p(L)e_0\|_2^2 \\
&= \|p(L)L^{1/2}e_0\|_2^2 \\
&\leq \|p(L)\|_2^2\|L^{1/2}e_0\|_2^2 \qquad \text{(Definition 2.12)} \\
&= \|p(L)\|_2^2\left\langle L^{1/2}e_0, L^{1/2}e_0\right\rangle \\
&= \|p(L)\|_2^2\langle Le_0, e_0\rangle \\
&= \|p(L)\|_2^2\|e_0\|_L^2 \\
&\leq \left(\max_{\lambda\in[m,M]}|p(\lambda)|\right)^2\|e_0\|_L^2 \qquad \text{(Theorem 3.6)}
\end{aligned}$$

Taking square roots and using the fact that $p \in P_j$ is arbitrary we deduce that

$$\inf_{p\in P_j}\|p(L)e_0\|_L \leq \left(\inf_{p\in P_j}\max_{\lambda\in[m,M]}|p(\lambda)|\right)\|e_0\|_L$$

as required. □

Therefore, in a Hilbert, space it is possible to estimate the convergence rate of CG by finding polynomials which, as j increases, have smaller and smaller magnitude on $[m, M]$ where m and M are strictly positive real numbers. Appealing to the scaled and shifted Chebyshev polynomials as we did in the finite dimensional case, we obtain the following result. Again, we mention that the rate of convergence given here could be made sharper.

Theorem 3.10 (Convergence of CG). *Let the* CG *iteration be applied to the problem $Lu = f$, where L is a bounded self-adjoint positive definite linear operator on a Hilbert space H. Suppose the spectrum of L is included inside the interval $[m, M]$, where m and M are strictly positive real numbers and with $m > 0$. Then* CG

converges in the L-norm and the errors of approximation satisfy

$$\|e_j\|_L \leq \left(\frac{M-m}{M+m}\right)^j \|e_0\|_L \tag{3.13}$$

Proof. If $m = M$ then m is the only spectral value of L. Then $L - mI$ has only 0 as spectral value and therefore $\langle (L - mI)x, x\rangle = 0$ for all $x \in H$. Since $L - mI$ is self-adjoint, it must be the 0-operator by [20, p. 404]. Hence $L = mI$ in which case f is an eigenvector of L and the *CG* algorithm converges in one step in agreement with Equation (3.7). If $m < M$, then the proof uses the scaled and shifted Chebyshev polynomials and is identical to the proof of Theorem 1.8. □

Once again we mention that the above bound for the rate of convergence could be tightened to obtain a similar equation to (1.7), namely

$$\|e_j\|_L \leq \frac{2}{\left[\left(\frac{\sqrt{M}+\sqrt{m}}{\sqrt{M}-\sqrt{m}}\right)^j + \left(\frac{\sqrt{M}+\sqrt{m}}{\sqrt{M}-\sqrt{m}}\right)^{-j}\right]} \|e_0\|_L. \tag{3.14}$$

Hence, the *CG* iteration converges in any Hilbert, space and the bound for its rate of convergence is the same as in the finite dimensional case.

Although we do not claim originality, we point out that the proof of Lemma 3.1 given here, one of the cornerstones of the proof of *CG*, is our own and we are not aware of the idea behind it being written anywhere else. A different proof of convergence of this result, which uses the Spectral Theorem, is given in the book by Daniel [10, pp. 117–121]. Our proof has the merit of being simpler, since it does not use the Spectral Theorem; in particular, Theorem 3.7 and Theorem 3.8, which we have used to prove Lemma 3.1, do not require the Spectral Theorem in their proofs.

Corollary 3.1. *Under the hypotheses of the previous theorem, the CG iterates also converge to the solution of the equation $Lu = f$ in the 2-norm, that is, in the original*

norm of the Hilbert space. In fact,

$$\|e_j\| \leq \sqrt{\frac{M}{m}} \left(\frac{M-m}{M+m}\right)^j \|e_0\|. \tag{3.15}$$

Proof. From Theorem 3.2 we see that for all $x \in H$,

$$m \langle x, x \rangle \leq \langle Lx, x \rangle \leq M \langle x, x \rangle .$$

Hence,

$$\sqrt{m}\|x\| \leq \|x\|_L \leq \sqrt{M}\|x\|$$

implying that the two norms are equivalent as $0 < m < M$. Since the *CG* iterates converge in the L-norm, they must also converge in the 2-norm. To establish (3.15), we use (3.13) and the above equivalence of norms to obtain

$$\sqrt{m}\|e_j\| \leq \|e_j\|_L \leq \left(\frac{M-m}{M+m}\right)^j \|e_0\|_L \leq \left(\frac{M-m}{M+m}\right)^j \|e_0\|\sqrt{M}.$$

Dividing by $\sqrt{m}$ yields Equation (3.15). □

3.3 Examples. Integral Equations.

We now provide several examples of *CG* in action in various Hilbert spaces. The implementation of the algorithm is based on the Maple software for symbolic manipulation and is described in Appendix A. We also give, in Appendix B, the specific code used to generate the output for each of our examples.

Example 3.1. *Consider the Hilbert space $H := L^2[1, 2]$. Our first example is based on a very simple operator L on H, defined by*

$$Lx(t) := tx(t). \tag{3.16}$$

This operator was already encountered in Example 2.9. Solving linear equations with this L is easily done directly. Nevertheless we will solve an equation involving L by

CG, *since we believe that the results will be instructive. The above operator is self-adjoint because multiplication is associative and commutative, and t is real valued:*

$$\langle tx, y\rangle = \int_1^2 tx(t)\overline{y(t)}dt = \int_1^2 x(t)\overline{ty(t)}dt = \langle x, ty\rangle$$

for all $x, y \in H$. *Moreover, L is positive definite (Theorem 3.2 and Definition 3.3), because* $t \in [1, 2]$, *so that for x, with* $\|x\| = 1$, *we have*

$$\langle Lx, x\rangle = \int_1^2 t|x(t)|^2\, dt,$$

which takes values between 1 and 2. Note that in fact multiplication by any integrable bounded function with strictly positive infimum defines a linear self-adjoint positive definite operator on H.

Consider the sample equation $(Lu)(t) = \sin(t)$. *The true solution is just* $\hat{u} = sin(t)/t$ *and is represented by the solid curve in Figure 3.1. This equation could be approximately solved by discretizing it and obtaining a system* $Au_d = \sin(t_d)$ *where* u_d *and* t_d *are just the discrete versions of u and t respectively. In that case A is just a diagonal matrix with the vector* t_d *in its diagonal. We now solve this equation by a direct and symbolic version of* CG. *The results, shown in Figure 3.1 and Table 3.1, compare the exact solution to the first six approximations.*

We have already seen in Example 2.9 that this operator has no eigenvalues, hence its point spectrum is empty, but nevertheless has a spectral value at each point in $[1, 2]$. *It is interesting to observe that the matrix approximation A of L also has a spectrum inside* $[1, 2]$ *becoming denser as the dimension of A increases.*

We now look at more involved examples taken from linear integral equations on the real space $L^2[a, b]$, where the unknown function is part of the integrand. Integral equations are widely used in applications and are intimately connected to differential equations, including boundary and initial value problems. A differential equation can

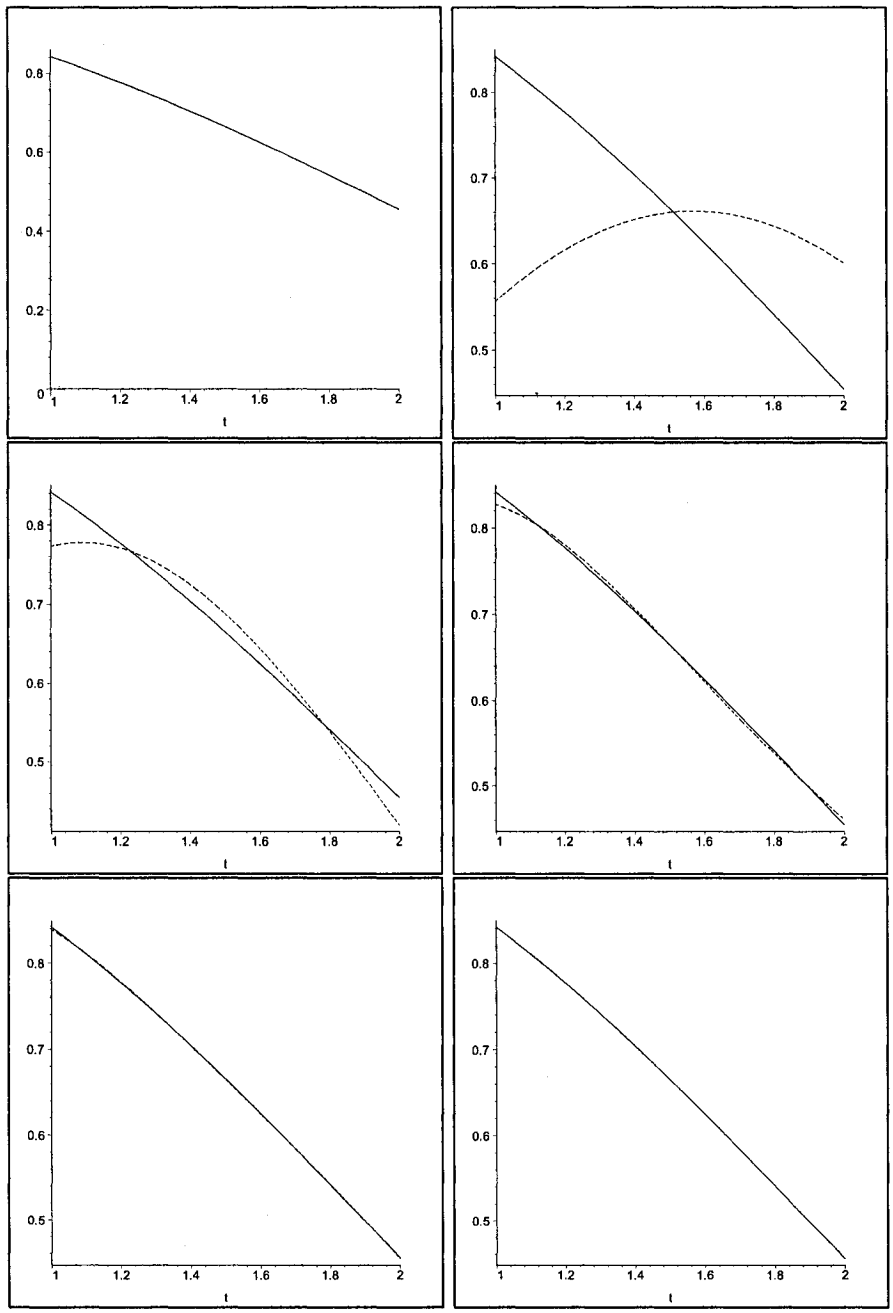

Figure 3.1: First 6 iterations of *CG* for Example 3.1.

Iteration	Approximation	$L-$norm of error
1	0	0.7928145140
2	$0.6615537583\sin(t)$	0.1490781938
3	$-0.4570617629\,t\sin(t) + 1.375913607\sin(t)$	0.02673638316
4	$0.3146365087\,t^2\sin(t) - 1.418952428\,t\sin(t)$ $+2.087319268\sin(t)$	0.004670521984
5	$-0.2161857918\,t^3\sin(t) + 1.299201009\,t^2\sin(t)$ $-2.882511937\,t\sin(t) + 2.796137112\sin(t)$	0.0008088553023
6	$0.1484540471\,t^4\sin(t) - 1.114828041\,t^3\sin(t)$ $+3.308148013\,t^2\sin(t) - 4.846514976\,t\sin(t)$ $+3.504097478\sin(t)$	0.0001395782934

Table 3.1: First 6 iterations of *CG* for Example 3.1.

often be transformed into an appropriate integral equation [23]. In general, integral operators tend to be better behaved than differential operators partly because they are smoothing operators, but mainly because they very often are bounded operators. The latter property makes them suitable for *CG*. Furthermore, integral equations have the advantage that no boundary or initial conditions are needed for the unknown function u of the equation $Lu = f$.

We focus our discussion on one particular class of integral equations called *Fredholm equations of the second kind.* Consider $L^2[a,b]$ and let $k : [a,b] \times [a,b] \to \mathbb{R}$. We can define an integral operator K by

$$(K\phi)(t) = \int_a^b k(s,t)\phi(s)ds. \tag{3.17}$$

A Fredholm equation of the second kind has the form

$$u(t) - \lambda \int_a^b k(s,t)u(s)ds = f(t). \tag{3.18}$$

The function $k(s,t)$ is called the *kernel* of the equation or the *kernel* of the integral operator. Using operator notation, we can rewrite the above equation as $Lu = f$ where $L = (I - \lambda K)$. In the following examples we only consider the case $\lambda \in \mathbb{R}$. The term *Fredholm* refers to the fact that the integral is definite, that is, with fixed endpoints. An equation where the endpoints of the integral depend on t is called a *Volterra integral equation.* We point out that a *Volterra* equation can also be written as a *Fredholm* equation. For example the Volterra equation

$$u(t) - \lambda \int_a^t k_1(s,t)u(s)ds = f(t) \tag{3.19}$$

can be rewritten as the Fredholm equation

$$u(t) - \lambda \int_a^b k_2(s,t)u(s)ds = f(t) \tag{3.20}$$

where $k_2(s,t) = k_1(s,t)$ for $0 < s < t$ and is 0 otherwise. The *second kind* in (3.18) refers to the fact that the unknown is present outside the integral sign, standing alone, as well as inside. A Fredholm equation of the first kind is similar to 3.18, except that the unknown is present only under the integral sign. The variety of possible kernels allow Fredholm equations of the first and second kind to be fairly general despite their restricted appearance. In other words, many differential and integral equations can be reduced to the form of (3.18).

We now gather some information about the operator K justifying whether we can apply *CG* to a given equation of type (3.18).

Proposition 3.2. *The integral operator K from Equation (3.18) is bounded, provided its kernel $k \in L^2([a,b] \times [a,b])$. In that case*

$$\|K\|_2 \leq \int_a^b \int_a^b |k(s,t)|^2 ds\, dt \tag{3.21}$$

Proof. Also see [16, p. 17]. For any $\phi \in L^2[a,b]$ we have, by the Cauchy-Schwarz

inequality (Proposition 2.1),

$$\begin{aligned}\|K\phi\|^2 &= \int_a^b \left(\int_a^b |k(s,t)\phi(s)|ds \right)^2 dt \\ &= \int_a^b \left(\int_a^b |k(s,t)|^2 ds \int_a^b |\phi(s)|^2 ds \right) dt \\ &= \|\phi\|^2 \int_a^b \int_a^b |k(s,t)|^2 ds\, dt\end{aligned}$$

yielding the required result. □

Proposition 3.3. *The integral operator K from Equation (3.17) is self-adjoint provided it is bounded and its kernel k satisfies*

$$k(s,t) = \overline{k(t,s)} \tag{3.22}$$

for almost all $s, t \in [a,b]$.

Proof. For the proof of this, we also refer the reader to [23, pp.104–105] and [16, p. 69]. First, K needs to be bounded simply by definition of a self-adjoint operator. Now, for any $\phi, \psi \in L^2[a,b]$,

$$\begin{aligned}\langle K\phi, \psi \rangle &= \int_a^b \left(\int_a^b k(s,t)\phi(s)ds \right) \overline{\psi(t)}dt \\ &= \int_a^b \int_a^b k(s,t)\phi(s)\overline{\psi(t)}ds\, dt \\ &= \int_a^b \int_a^b k(s,t)\phi(s)\overline{\psi(t)}dt\, ds \\ &= \int_a^b \left(\phi(s) \int_a^b \overline{k(t,s)\psi(t)}dt \right) ds \\ &= \int_a^b \left(\phi(s) \overline{\int_a^b k(t,s)\psi(t)dt} \right) ds \\ &= \langle \phi, K\psi \rangle\end{aligned}$$

The change of order of integration is justified by *Fubini's* theorem, since the kernel belongs to $L^2([a,b] \times [a,b])$. □

Iteration	Approximation	$L-$norm of error
1	0	2.090348697
2	2.751938394	0.2163861243
3	$1.399380674\cos(t) + 1.399380674\sin(t)$ $+1.000000000$	0.0000000000

Table 3.2: First 3 iterations of *CG* for example 3.2

For real kernels, the above condition reduces to *symmetry*, that is, $k(s,t) = k(t,s)$. We now solve some Fredholm integral equations of the second kind using *CG*.

Example 3.2. *Consider the equation*

$$u(t) - \frac{1}{2}\int_0^{\frac{\pi}{2}} \sin(s+t)u(s)ds = 1 \tag{3.23}$$

A direct computation shows that $\|K\| < \pi^2/8 + 1/2 < 1.8$ *(Proposition 3.2); hence* $\sigma(K) \subseteq [-1.8, 1.8]$ *by Theorem 3.4. Thus by Proposition 2.10,* $\sigma(I-K/2) \subseteq [0.1, 1.9]$*; the exact spectral bounds may be difficult to determine. Moreover, the kernel of the equation is symmetric and hence the operator* $L = I - \lambda K$ *is self-adjoint positive definite. The results of applying 3 iterations of* CG *are given in Table 3.2 and Figure 3.2. Each approximation is plotted as a dotted curve together with the exact solution as a solid curve. At the first iteration, only the solid curve appears because the approximation coincides with the* x*-axis. At the last iteration, only one solid curve appears because the approximation and the true solution cannot be distinguished. The exact solution of the equation is given by:*

$$u(t) = 1 + \frac{4}{6-\pi}\sin t + \frac{4}{6-\pi}\cos t. \tag{3.24}$$

From Table 3.2 we can see that the $L-$norm of the error is 0, to 10 significant digits! In fact, the exact solution can also be found, in symbolic form, by Maple at

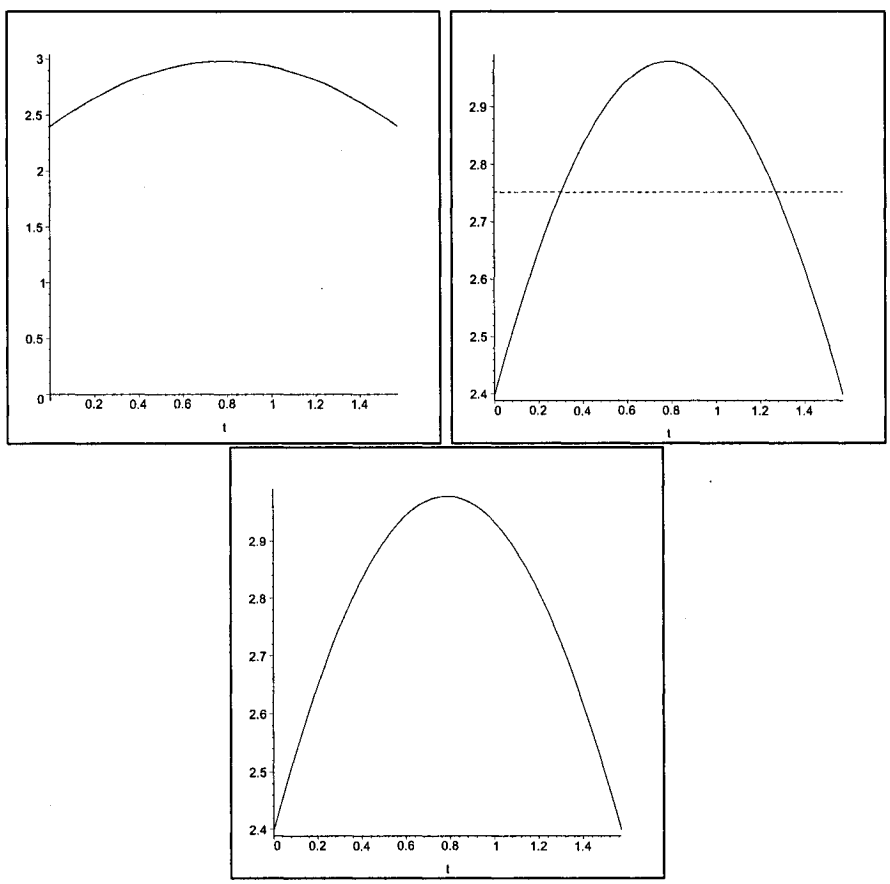

Figure 3.2: First 3 iterations of CG for example 3.2

the third iteration using a purely symbolic version of the *CG*, where all constants are kept in symbolic form as opposed to decimal form. Exact computations such as this are possible for simple equations but are generally very time consuming and impractical for more complicated problems because symbolic expressions such as rational numbers, roots and constants increase quickly in size with the number of iterations. For a discussion of this issue, see Appendix A.

Why did *CG* converge in only two iterations in the above example? The reason behind this behavior is the type of kernel used in the integral operator: $\sin(s+t) = \sin(s)\cos(t) + \cos(s)\sin(t)$. In general, if a kernel can be expressed as

$$k(s,t) = \sum_{j=1}^{n} a_j(s)b_j(t) \tag{3.25}$$

then Equation (3.18) can be rewritten as

$$u(t) - \lambda \int_a^b \left(\sum_{j=1}^{n} a_j(s)b_j(t) \right) u(s)ds = f(t) \tag{3.26}$$

which is equivalent to

$$u(t) = f(t) + \sum_{j=1}^{n} \alpha_j b_j(t) \tag{3.27}$$

where $\alpha_j = \lambda \int_a^b a_j(s)u(s)ds$. This means that the solution is just a linear combination of $f(t), b_1(t) \ldots, b_n(t)$ and hence, that the problem is finite dimensional and could be solved simply by a system of linear equations with unknowns $\alpha_1, \ldots, \alpha_n$; see [23, pp.57–60]. A kernel with the above property is called *degenerate* or *separable*. It is now easy to understand why the solution $u(t)$ of Example 3.2 has the form given by (3.24). The equations in our remaining examples possess non-degenerate kernels.

Notice that in the previous example, positive definiteness of $L = I - \lambda K$ was shown using the fact that $\|\lambda K\| < 1$. This requirement simplifies the verification of

positive definiteness of L but is rather limiting. We next demonstrate a technique used to show directly that an integral operator K is positive, resulting in a wider range of possible values λ for which Equation (3.18) is solvable by *CG*. To simplify our task, we first state two technical propositions.

Proposition 3.4. *Let $\phi \in L^2[a,b]$ then*

$$\int_a^b \int_t^b \phi(s,t)ds\,dt = \int_a^b \int_a^s \phi(s,t)dt\,ds. \tag{3.28}$$

Proof. This formula is just a change of the order of integration in a triangular region. Let

$$h(s,t) = \begin{cases} 1 & ,s \geq t, \\ 0 & ,s < t. \end{cases}$$

Then for $s,t \in [a,b]$, Fubini's theorem yields,

$$\begin{aligned} \int_a^b \int_t^b \phi(s,t)ds\,dt &= \int_a^b \int_a^b h(s,t)\phi(s,t)ds\,dt \\ &= \int_a^b \int_a^b h(s,t)\phi(s,t)dt\,ds \\ &= \int_a^b \int_a^s \phi(s,t)dt\,ds. \end{aligned}$$

□

Proposition 3.5. *Let T be a bounded operator on $L^2[a,b]$ defined by*

$$(T\phi)(t) = \int_t^b \phi(s)ds, \tag{3.29}$$

then its adjoint T^* is given by

$$(T^*\phi)(t) = \int_a^t \overline{\phi(s)}ds. \tag{3.30}$$

Proof. Although our examples are in real $L^2[a,b]$, we do the proof in the general complex case. Let ϕ, ψ be members of the complex Hilbert space $L^2[a,b]$. Then, by

an application of the previous proposition, we obtain

$$\begin{aligned}
\langle T\phi, \psi \rangle &= \int_a^b \left(\int_t^b \phi(s) ds \right) \overline{\psi(t)}\, dt \\
&= \int_a^b \int_t^b \phi(s) \overline{\psi(t)} ds\, dt \\
&= \int_a^b \int_a^s \phi(s) \overline{\psi(t)} dt\, ds \\
&= \int_a^b \phi(s) \left(\overline{\int_a^s \psi(t) dt} \right) ds \\
&= \langle \phi, T^* \psi \rangle .
\end{aligned}$$

It readily follows that

$$\left(T^*\phi\right)(t) = \int_a^t \overline{\phi(s)} ds$$

as required. □

Proposition 3.6. *The Fredholm operator K on $L^2[0,1]$ defined by*

$$\left(K\phi\right)(t) = \int_0^1 \min(s,t)\phi(s) ds \tag{3.31}$$

is positive.

Proof. This proof is a modified version of the one given in [23, p. 169]. It is based on the observation that for a bounded operator T on $L^2[0,1]$, we have $\langle T^*T\phi, \phi \rangle = \langle T\phi, T\phi \rangle \geq 0$ so the operator T^*T is positive. The idea of the proof is to factor K into T^*T. Since the latter is given by a double integral, we rewrite the simple integrals as double integrals; after splitting and regrouping the terms, we obtain the desired

form. Here are the details:

$$\begin{aligned}
(K\phi)(t) &= \int_0^1 \min(s,t)\phi(s)ds \\
&= \int_0^t s\phi(s)ds + \int_t^1 t\phi(s)ds \\
&= \int_0^t \left(\int_0^s dr\right)\phi(s)ds + \int_t^1 \left(\int_0^t dr\right)\phi(s)ds \\
&= \int_0^t \int_0^s \phi(s)dr\,ds + \int_0^t \int_t^1 \phi(s)dr\,ds \\
&= \int_0^t \int_r^t \phi(s)ds\,dr + \int_0^t \int_t^1 \phi(s)ds\,dr \\
&= \int_0^t \int_r^1 \phi(s)ds\,dr \\
&= (T^*T\phi)(t),
\end{aligned}$$

where

$$(T\phi)(r) = \int_r^1 \phi(s)ds \quad \textit{and} \quad (T^*\psi)(t) = \int_0^t \phi(r)dr$$

from Proposition 3.5. The change of variables is justified by Fubini's theorem and Proposition 3.4 □

The approach used in the proof of Proposition 3.6 is very useful. Together with Proposition 2.7, one can use it to construct a considerable collection of self-adjoint positive operators which appear in various integral kernels. Positive operators K on $L^2[0,1]$ found using the above ideas include those generated by the kernels $1 - \max(s,t)$, $\frac{1}{2} - |t-s|$, $-\ln|t-s|$, $\frac{1}{t+s}$, $e^{-|t-s|}$ and others. For more information about positive integral operators, we refer the reader to Chapter 6 in the book by Porter and Stirling [23].

Although it may be difficult to obtain the exact spectral bounds m_K and M_K for a positive operator K on $L^2[a,b]$, we know that $L = I - \lambda K$ is positive definite for any $\lambda \in (-\infty, \frac{1}{\|K\|})$ (Proposition 2.10) and hence suitable for *CG*.

Example 3.3. *Consider the equation*

$$u(t) + 5\int_0^1 \min(s,t)u(s)ds = f(t) \tag{3.32}$$

where

$$f(t) = \frac{10696}{10201}e^t\cos(10\,t) - \frac{100}{10201}e^t\sin(10\,t) - \frac{495}{10201} + \frac{5e}{101}t\cos(10) + \frac{50e}{101}t\sin(10)\,.$$

The reason we have chosen this complicated value of f is that we have started with the solution $u(t) = e^t\cos(10t)$ to which we have applied the operator L and have obtained $f(t)$. Our purpose in this paper is not to solve equations with unknown solutions, but to demonstrate to the reader that CG *works in a Hilbert space other than $\mathbb{R}^n$ or $\mathbb{C}^n$, and could be practically implemented to solve problems of interest. The integral operator K of equation (3.32) is non-degenerate since there is no way to write $\min(s,t)$ in the form of (3.25). Moreover, K is bounded, self-adjoint and is positive by Proposition 3.6. Therefore $L = I + 5K$ is self adjoint positive definite with a lower bound $m_L \geq 1$. Using Proposition 3.2, we can also deduce that $\|K\|_2 \leq 8/3$. Therefore, the spectrum of L satisfies $\sigma(L) \in [1, \frac{8}{3}]$ giving an indication of an upper bound for the rate of convergence of* CG *applied to (3.32). To ease Maple computations, we rewrite the equation in the equivalent form*

$$u(t) + 5\int_0^t s\,u(s)ds + 5t\int_t^1 u(s)ds = f(t). \tag{3.33}$$

The results after applying 6 iterations of CG *to the problem are given in Table 3.3 and are also shown in Figure 3.3. The exact result is plotted as a solid curve together with the approximation as a dotted curve for each iteration. At the first iteration, the approximation is 0 and coincides with the x-axis. At the fourth iteration, the approximation and solution curves cannot be distinguished and only one solid curve appears. In Figure 3.4, we also give a closer look at the final two approximations by providing a plot of the errors.*

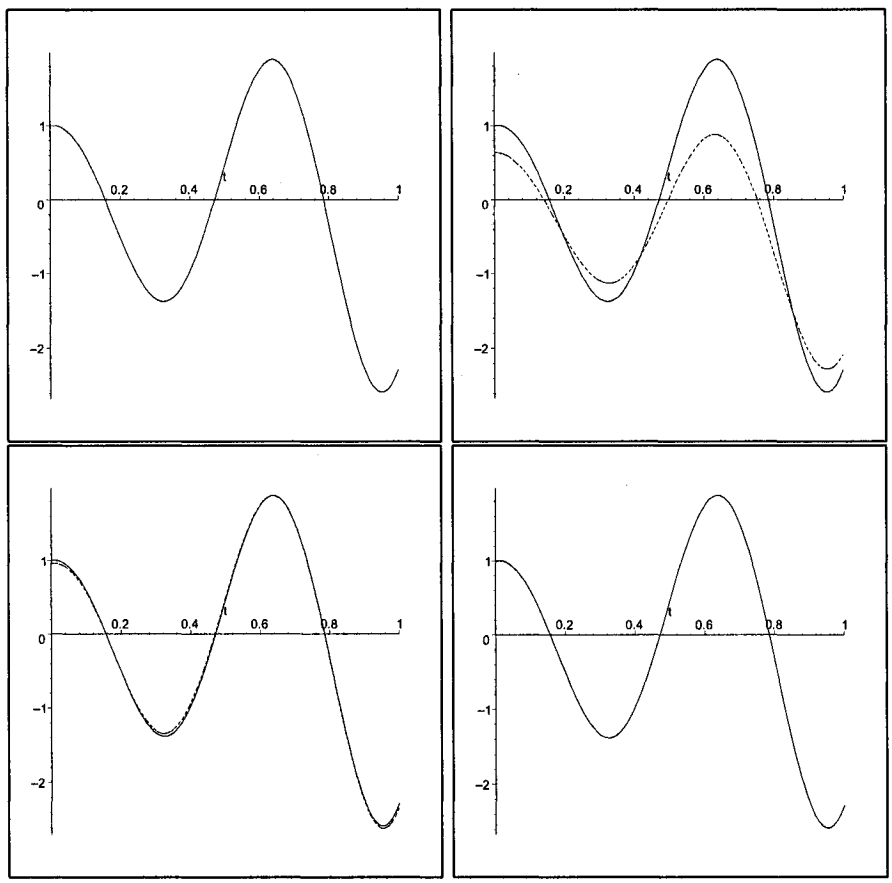

Figure 3.3: First 4 iterations of *CG* for Example 3.3.

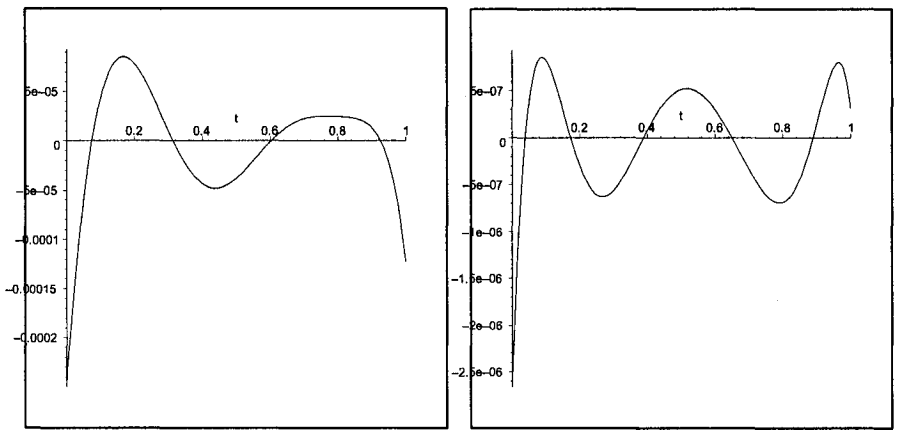

Figure 3.4: Error of the 5^{th} and 6^{th} approximations for Example 3.3.

In Table 3.3, it is interesting to observe that the coefficient of $e^t \cos(10t)$ *seems to approach 1 as the iterations increase, while the rest of the expressions appear to approach 0.*

Our next example is interesting in that it takes place on the infinite interval $(-\infty, \infty)$. Before we proceed, we will need the following result taken verbatim from [23, p. 333] describing the spectrum of certain types of integral operators on $L^2(-\infty, \infty)$.

Proposition 3.7. *Suppose that* $k \in L^1(-\infty, \infty)$. *Then the formula* $(K\phi)(t) = \int_{-\infty}^{\infty} k(t-s)\phi(s)\,ds$ *defines a bounded operator* K *from* $L^2(-\infty, \infty)$ *to itself whose spectrum is* $\overline{\{\hat{k}(t) : t \in \mathbb{R}\}}$ *where* $\hat{k}(t) = \int_{-\infty}^{\infty} k(s)e^{its}\,ds$ *and which is self-adjoint if and only if* $\hat{k}(t)$ *is real-valued.*

Proof. [23, pp. 332-333]. □

Example 3.4. *Consider the equation*

$$u(t) + \int_{-\infty}^{\infty} \frac{e^{-(t-s)^2}}{\sqrt{\pi}} u(s)ds = e^{-t^2}\left(t^3 - \frac{3}{4}t\right) + \frac{\sqrt{2}}{16}e^{-\frac{t^2}{2}}t^3. \tag{3.34}$$

Iteration	Approximation	$L-$norm of error
1	0	1.425251877
2	$0.6688991175e^t \cos(10t) - 0.006253731467e^t \sin(10t)$ $-0.03095597077 - 0.5390575344\,t$	0.6106518948
3	$0.9872210796e^t \cos(10t) - 0.006051478525e^t \sin(10t)$ $-0.03071256419 + 0.1970752527\,t - 0.03750840181\,t^2$ $-0.2177198357t^3 + 1 \times 10^{-10}e^t \sin(10t)t$	0.03021074563
4	$0.9996400504e^t \cos(10t) - 0.000670460205e^t \sin(10t)$ $-0.00473159760 + 0.020717440\,t - 0.07157330127\,t^2$ $+0.1121565701\,t^3 - 0.01351953711\,t^4 - 0.0470849932\,t^5$ $+2.0 \times 10^{-10}e^t \sin(10t)t - 1.0 \times 10^{-10}e^t \cos(10t)t$ $-2.0 \times 10^{-21}e^t \cos(10t)t^2 - 2.0e \times 10^{-22}e^t \sin(10t)t^2$	0.002235888657
5	omitted	$5.319284173 \times 10^{-5}$
6	omitted	$5.445860314 \times 10^{-7}$

Table 3.3: First 6 iteration of CG for Example 3.3.

The kernel of the integral operator K is the well-known Gaussian kernel. *Since it is of the form $k(t-s)$ where $k(t) = \frac{e^{-t^2}}{\sqrt{\pi}}$ and since $k \in L^1(-\infty, \infty)$, the previous proposition applies. A direct calculation shows that $\hat{k}(t) = e^{-\frac{t^2}{4}}$ indicating that K is self-adjoint and that $\sigma(K) \subseteq [0,1]$. It follows that the operator $L = (I+K)$ of (3.34) is self-adjoint positive definite with $\sigma(L) \subseteq [1,2]$. Observe that* CG *is guaranteed to work for the above equation because the right-hand-side $f(t) \in L^2(-\infty, \infty)$. This condition on f is necessary since the theory of* CG *is based on the assumption that L is an operator mapping a Hilbert space, $L^2(-\infty, \infty)$in this case, to itself. If $f \notin L^2(-\infty, \infty)$, Equation (3.34) may still have a solution but* CG *is not guaranteed to converge to it. The true solution of (3.34) is given by*

$$u(t) = e^{-t^2}\left(t^3 - \frac{3}{4}t\right). \tag{3.35}$$

Applying six iterations of CG *to the equation yields the results given in Table 3.4. Also, Figure 3.5 illustrates the first four iterates of* CG *together with the solution. As before, the true solution is plotted as a solid curve together with the approximation as a dotted curve. At the first iteration the approximation is not seen since it coincides with the x-axis, whereas at the fourth iteration, the two curves almost coincide so that only one solid curve appears. We also provide a plot of the errors for the 5^{th} and 6^{th} approximation on Figure 3.6.*

Our last example demonstrates that CG may fail to work if one of the assumptions of Theorem 3.10 is not satisfied. Here we consider an operator which is unbounded. For such operators, considerations about domain become of prime importance as the following theorem shows.

Theorem 3.11 (Hellinger-Toeplitz). *Let H be a Hilbert space. Let L be a linear operator defined on all of H. If L satisfies*

$$\langle Lx, y\rangle = \langle x, Ly\rangle \tag{3.36}$$

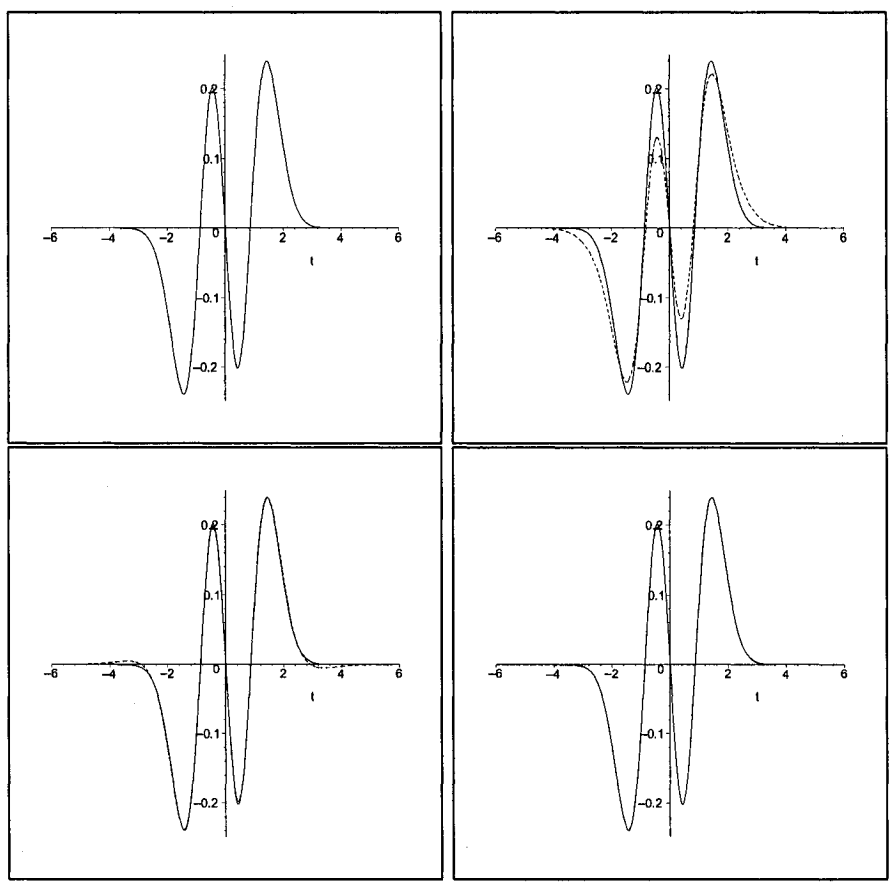

Figure 3.5: First 4 *CG* iterations for Example 3.4.

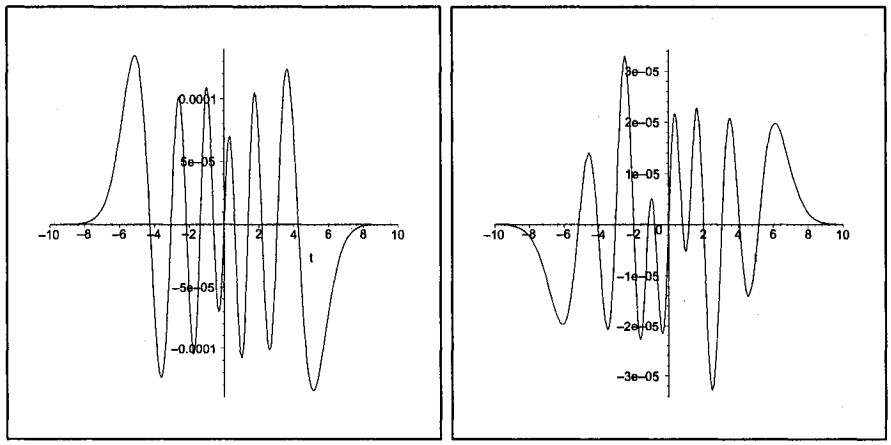

Figure 3.6: Error of the 5^{th} and 6^{th} approximations for Example 3.4.

Iteration	Approximation	$L-$norm of error
1	0	0.3957064512
2	$0.6646596122\,t^3e^{-t^2} - 0.4984947092\,te^{-t^2}$ $0.05874816485\,t^3e^{-\frac{t^2}{2}} - 1.0 \times 10^{-10}\,te^{-\frac{t^2}{2}}$	0.09629803217
3	$0.9470292743\,t^3e^{-t^2} - 0.7102719558\,te^{-t^2}$ $+0.04145784593\,t^3e^{-\frac{t^2}{2}} - 8.0 \times 10^{-11}\,te^{-\frac{t^2}{2}}$ $-0.01022096631\,t^3e^{-\frac{t^2}{3}} - 0.02299717419\,te^{-\frac{t^2}{3}}$	0.01181583031
4	$0.9862350252\,t^3e^{-t^2} - 0.7396762689\,te^{-t^2}$ $+0.01685400978\,t^3e^{-\frac{t^2}{2}} + 9.0 \times 10^{-11}\,te^{-\frac{t^2}{2}}$ $-0.01024386309\,t^3e^{-\frac{t^2}{3}} - 0.02304869194\,te^{-\frac{t^2}{3}}$ $+0.002472621909\,t^3e^{-\frac{t^2}{4}} + 0.01483573145\,te^{-\frac{t^2}{4}}$	0.002443540951
5	omitted	$3.495969179 \times 10^{-4}$
6	omitted	$6.746692868 \times 10^{-5}$

Table 3.4: First 6 iteration of CG for Example 3.4.

for all $x, y \in H$, then L is bounded.

Proof. [19, pp. 525-526] □

Corollary 3.2. *An unbounded linear operator L satisfying (3.36) cannot be defined on all of H.*

From the corollary, it follows that an unbounded operator L is usually defined together with a domain $\mathcal{D}(L) \subset H$. In this case, we replace the usual definition of self-adjointness by the more general definition of *symmetry.*

Definition 3.12. Let H be a Hilbert space and let $L : \mathcal{D}(L) \subseteq H \to H$ be a linear operator (bounded or not) satisfying (3.36) for all $x, y \in \mathcal{D}(L)$. Then L is called *symmetric* on $\mathcal{D}(L)$.

Notice that $\langle Lx, y\rangle$ is defined on $\mathcal{D}(L) \times H$ and $\langle x, Ly\rangle$ is defined on $H \times \mathcal{D}(L)$. Since the common domain of these inner products is $\mathcal{D}(L) \times \mathcal{D}(L)$, the notion of symmetry requires equality only on that common domain, disregarding the rest of H. It is conceivable therefore for L to be symmetric on one domain, but not on another. In particular, every linear operator is symmetric on $\{0\} \subseteq H$. In our case, we take $\mathcal{D}(L)$ as the largest possible domain of L, that is, $\mathcal{D}(L) = \{x \in H : \ Lx \in H\}$.

Example 3.5. *Let $H = L^2[0,1]$ and let $L : \mathcal{D}(L) \to H$ be defined by*

$$Lx(t) = \frac{1}{t}x(t), \tag{3.37}$$

where

$$\mathcal{D}(L) := \{x(t) \in H : \frac{x(t)}{t} \in H\}.$$

Here the point spectrum of L is empty (see Definition 2.18) because L has no eigenvalues, that is, no non-zero function $x(t) \in \mathcal{D}(L)$ satisfies $\frac{1}{t}x(t) = x(t)$. Also, the

residual spectrum is empty because L_λ^{-1} *is just defined by multiplication by* $\left(\frac{1}{t}-\lambda\right)^{-1} = \frac{t}{1-\lambda t}$ *and has a dense domain in* H. *In particular, the set*

$$\{y(t) \in H : \; y(t) = \textit{polynomial in } t, \; \textit{multiple of } (1-\lambda t)\}$$

is part of the domain of L_λ^{-1} *and is dense in* H *by a theorem of Müntz[1, p.43] for any* $\lambda \in \mathbb{C}$. *Finally, the continuous spectrum of* L *consists of the interval* $[1,\infty]$ *because the resolvent* L_λ^{-1} *is unbounded for* $\lambda \in [1,\infty]$ *and is bounded for any other* $\lambda \in \mathbb{C}$. *Therefore the spectrum of* L *is given by* $\sigma(L) = [1,\infty]$. *Moreover,* L *is symmetric on* $\mathcal{D}(L)$ *by associativity and commutativity of multiplication and since* t *is real valued. Consider the equation,*

$$Lu(t) = t^2. \tag{3.38}$$

This equation is consistent in H *and the true solution can be found directly, by multiplication by* t. *It is given by* $u(t) = t^3$ *and is unique up to a set of Lebesgue-measure zero. Applying six iterations of* CG *to this equation yields the results shown in Figure 3.5 and Table 3.5. The true solution is plotted as a solid curve together with the approximation as a dotted curve. At the first iteration, the approximation is not seen since it coincides with the* x*-axis. From Figure 3.5 and Table 3.5, we see that* CG *fails to converge to the solution of (3.38). A close examination of the successive Krylov subspaces reveals that, in fact,* $K_4 = span\{t^2, L(t^2), L^2(t^2), L^3(t^2)\}$ *does not lie in* H *since* $L^3(t^2) = \frac{1}{t} \notin H$. *It is worth pointing out that the* Maple *implementation of* CG, *given in the Appendix A, constructs* $L^3(t^2)$ *nevertheless, but that the step length in the direction of* $L^3(t^2)$ *is calculated as 0 since any other value of the step length would produce an infinite error in the L-norm.*

3.4 CG in practice

Although in theory *CG* works for bounded self-adjoint positive definite operators, its practical implementation and use is a different matter. Solving $Lu = f$ by *CG*

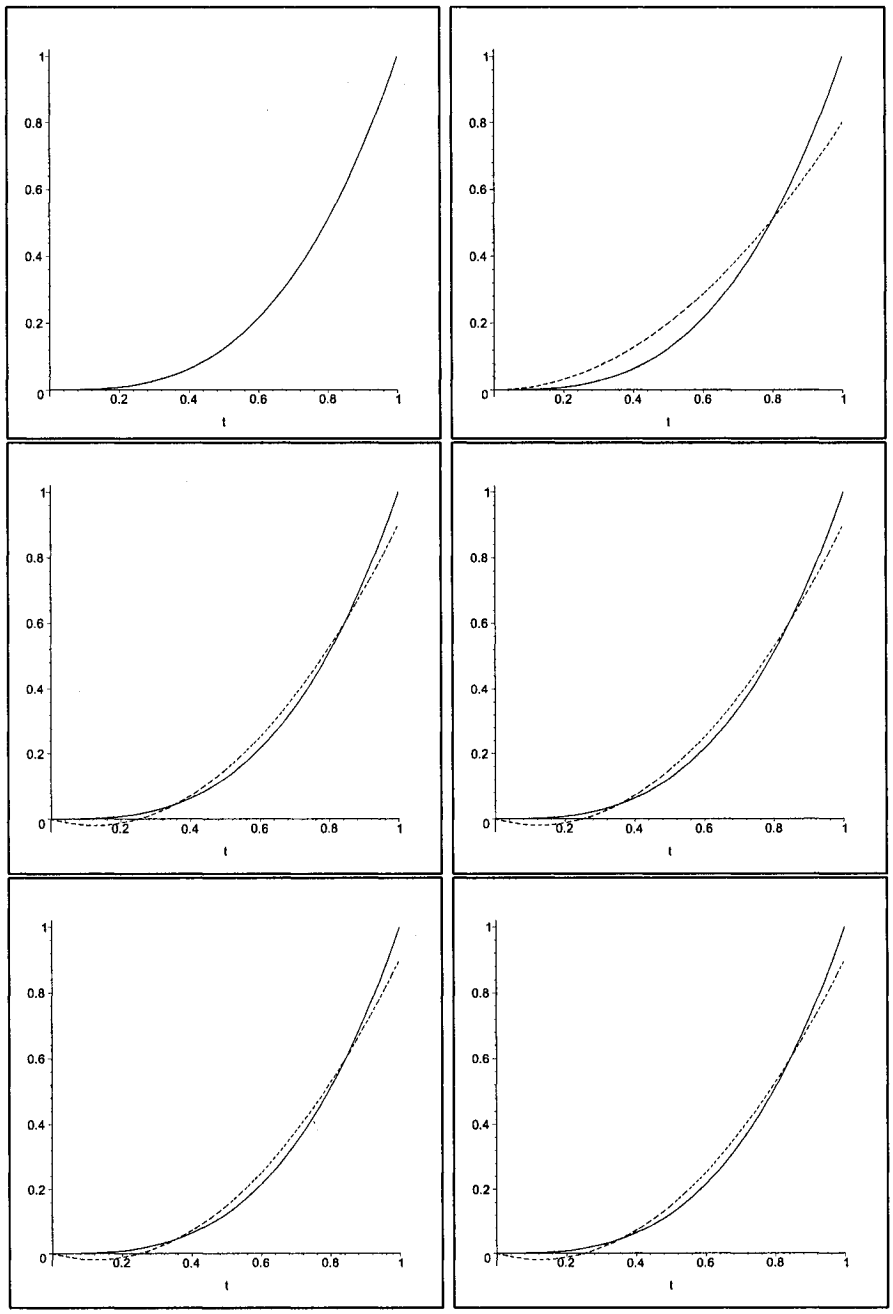

Figure 3.7: First 6 iterations of *CG* for Example 3.5.

Iteration	Approximation	$L-$norm of error
1	0	0.4082482905
2	$0.8000000000\,t^2$	0.08164965809
3	$1.200000000\,t^2 - 0.2999999999\,t$	0.04082482905
4	$1.200000000\,t^2 - 0.2999999999\,t$	0.04082482905
5	$1.200000000\,t^2 - 0.2999999999\,t$	0.04082482905
6	$1.200000000\,t^2 - 0.2999999999\,t$	0.04082482905

Table 3.5: First 6 iteration of CG for Example 3.5.

assumes that $Lf, L^2f, L^3f, \ldots$ are easy to evaluate but this is not always the case. In our examples of integral equations, we used expressions which can be readily calculated by *Maple*. Many integral operators are difficult to evaluate, however, making the direct application of CG inefficient. One approach to overcome this difficulty is to solve the perturbed problem $L\tilde{u} = \tilde{f}$ where $L^k\tilde{f}$ is easy to evaluate and $\tilde{f}$ is sufficiently close to f; then $\tilde{u}$ will be close to u. The idea is precisely stated as follows. Let

$$L : H \to H$$

be a bounded self-adjoint positive definite operator on a Hilbert space H with spectrum

$$\sigma(L) \subseteq [m, M]$$

where $m > 0$. Then L^{-1} exists, is self-adjoint positive definite and has spectrum

$$\sigma(L^{-1}) \subseteq \left[\frac{1}{M}, \frac{1}{m}\right].$$

Thus,

$$\|L^{-1}\| \leq \frac{1}{m}$$

from Theorem 3.9. Subtracting $Lu = f$ from $L\tilde{u} = \tilde{f}$ and applying L^{-1} on both sides, we obtain

$$\|\tilde{u} - u\| = \|L^{-1}(\tilde{f} - f)\| \leq \|L^{-1}\|\|\tilde{f} - f\| \leq \frac{\|\tilde{f} - f\|}{m}.$$

giving an exact idea of how close $\tilde{u}$ will be to u with respect to the initial perturbation. We end this section by listing some characteristics of continuous *CG* and discrete *CG* applied to a given problem allowing the reader to compare the advantages and disadvantages of each.

Continuous CG

- simple to use (advantage): avoids the discretization step.
- direct application to the equation of interest (advantage).
- restricted applicability (disadvantage): it is not always possible to evaluate $L^k f$.

Discrete CG

- indirect application to the equation of interest (disadvantage): the results depend on how well the discrete system approximates the original equation.
- wide applicability (advantage): it is always possible to evaluate $A^k b$.

Conclusion

In Chapter 1 of this thesis, we considered the CG method for solving systems of linear equations where the corresponding matrix is self-adjoint and positive definite. In Chapter 2, we saw that standard matrix theory largely extends to linear operators in an arbitrary Hilbert space H. In particular, we made precise the idea of a self-adjoint positive definite operator on H allowing us to extend the finite-dimensional CG-theory to H. Finally, we used examples to show that a direct version of CG can be successfully implemented to solve equations of interest, in particular, integral equations.

We believe that studying CG for $Lu = f$ on an arbitrary Hilbert space is important for two reasons:

1. It provides important insight for the behavior of CG applied to a discrete system corresponding to $Lu = f$.

2. It provides an alternative approach to problem solving.

The exposition of CG in this thesis is by no means exhaustive. We end by listing some questions which have not been discussed, but which are worth considering, and which we intend to study in the future:

- Can CG be applied to self-adjoint positive semi-definite operators? An affirmative answer would, in particular, give the possibility of solving certain Fredholm

equations of the first kind.

- Under what conditions, if any, can *CG* be applied to equations with unbounded operators such as differential equations?

- In a general Hilbert space, to what extent can we apply *GMRES* and other Krylov subspace methods to solve linear equations symbolically?

Appendix A

Conjugate Gradient for Self-Adjoint Positive Definite Operators

Description:

Approximates the solution to the linear system $Lu = f$ where L is a self-adjoint positive definite linear operator. More precisely, at step n, the algorithm finds the vector v in the Krylov space K_n such that the L-norm of $v - u$ is minimal. The algorithm is standard and we do not describe the exact mechanics behind it. These can be found in [26], [13, pp. 520-541], [27, pp. 293–302] and [11, pp. 307–319]. The specific code used here is based on the one given by Trefethen [27, p. 294] with the exception of the stopping criterion which was inspired by [13, p. 529]. We warn the reader that, to avoid redundant computation, we have used some dummy variables storing values which are needed more than once; this makes the code slightly harder to read. Finally, we mention that the code is primarily intended for linear equations in a function space. To solve linear systems in $\mathbb{R}^n$, one should add, for technical reasons, the Maple function `evalm()` on the right hand side of the assignments for

`u[i]`, `r` and `p` inside the main loop.

Input:

L: a user defined linear operator.

f: the right hand side of the equation; a real vector if *CG* is performed in $\mathbb{R}^n$; an expression in t or x otherwise.

N: maximum number of iterations.

tol: used as a second termination criterion . The iteration stops if $\|r\|_2^2 < tol$.

Local Variables:

u: a list of the successive approximations found by CG.

r: residual.

r_prev: residual at the previous step.

p: search direction.

alpha: step-length.

beta: improvement this step.

i: loop counter.

tempLp, **tempInnProd_r_prev**, **tempInnProd_r**: dummy variables used to avoid redundant computation by storing values needed more than once. Their names describe what they store.

User Defined Functions:

L: a linear operator, typically defined as a procedure.

InnProd: a user defined procedure calculating the inner product of two elements from the space of interest.

Output:

u[1..i]: a list of all approximations to the linear equation. If termination occurs before N steps, a warning message will appear indicating the number of steps performed.

Maple Code:

```
CG := proc(L, f, N, tol)
    Declaring the local variables
    local u, r, r_prev, p, alpha, beta, i, tempLp,
    tempInnProd_r_prev, tempInnProd_r;

    Initializing the variables
    i:=1;
    u:=[seq(0,i=1..N)];
    u[1]:=0;
    r:=f;
    r_prev:=f;
    p:=r;
    alpha:=0;
    beta:=0;
    tempInnProd_r:=InnProd(r,r);

    Main iteration. The function expand removes nested parentheses.
    while i<N and evalf(tempInnProd_r)>tol do
        i:=i+1;
        tempLp:=L(p);
        alpha:=tempInnProd_r/InnProd(p,tempLp);
        u[i]:=expand(u[i-1]+alpha*p);

        Relocation of certain values followed by redefinition of the freed variables.
        r_prev:=r;
        tempInnProd_r_prev:=tempInnProd_r;
```

```
      r:=r_prev-alpha*tempLp;
      tempInnProd_r:=InnProd(r,r);

      beta:=tempInnProd_r/tempInnProd_r_prev;
      p:=r+beta*p;
   end do;

   Output
   if i<N then print("Premature termination at iteration:", i) end if;
   u[1..i];
end proc:
```

Remarks:

This code evaluates the constants numerically if the input constants are given in this form; otherwise they are kept in symbolic form. We encourage the numerical evaluation of constants to speed up the computation and to avoid huge symbolic expressions such as rational numbers, roots and others. If one wishes to perform a symbolic manipulation of the constants, however, we suggest adding the function `simplify` inside the code wherever expressions are manipulated. For example, instead on writing `alpha:=tempInnProd_r/InnProd(p,tempLp)`, one should write `alpha:=simplify(tempInnProd_r/InnProd(p,tempLp))`.

Appendix B

Maple Code Used in the Examples

General Remarks:

The numerical precision in all our examples is 10 significant digits. We also point out that no output is given here; the output was collected and discussed in the main text. Finally, in the current context, the expression *f is evaluated numerically* means that all constants in the function f are expressed in decimal form.

Code for Example 1.1:

Here, we have used the slightly modified implementation of *CG* by adding `evalm()` on the right-hand-side of certain assignments as explained in the beginning of Appendix A. First, we need the following package containing the matrix routines multiply and band.

```
with(linalg);
```

Defining the usual inner product for vectors in $\mathbb{R}^n$.

Input: v1, v2 are lists of length n; in the example, they play the role of column vectors.

Output: numerical value of the real inner product of v1, v2

```
InnProd:=proc(v1,v2)
```

```
   multiply(v1,v2);;
end proc;
```

Creates the matrix from Example 1.1. First we create a banded matrix, then we change the required entries to obtain the form given in Example 1.1. Finally we include $(1/h^2)$ inside A.

```
A:=band([-1,2,-1],5):
A[1,1]:=1:
A[1,2]:=0:
A[2,1]:=0:
A[4,5]:=0:
A[5,4]:=0:
A[5,5]:=1:
A:=evalm(1/0.25^2*A):
```

Defines the operator corresponding to multiplication by A.

```
L:=proc(v)
   multiply(A,v);
end proc:
```

Creating b.

```
b:=[0.,0.75,3.,6.75,0.];
```

Solves the problem of Example 1.1 using CG. The resulting iterates are stored in the sequence 'Approx'.

```
Approx:=CG(L,b,5,0.000000001);
```

The first iterate is just 0. We create the graphs (pairs of points) for each of the following iterates; there is only three more since the *CG* gave a warning of premature termination.

```
graph2:=[seq([i*0.25,Approx[2][i+1]],i=0..4)];
graph3:=[seq([i*0.25,Approx[3][i+1]],i=0..4)];
graph4:=[seq([i*0.25,Approx[4][i+1]],i=0..4)];
```

Finally, we plot all iterates together with the true solution.

```
plot([graph2,graph3,graph4,-t^4+t],t=0..1,linestyle=[DOT,DOT,DOT,SOLID]);
```

Code for Example 3.1:

Defining the usual inner product for real-valued functions on $[1, 2]$.

Input: f and g are expressions in terms of t.

Output: Numerical value of the inner product of f and g.

Remark: The integral is first evaluated symbolically, if possible, and then numerically. We can force Maple to perform direct numerical integration by writing `evalf(Int(f*g,t=1..2))`.

```
InnProd:=proc(f,g)
   evalf((int(f*g,t=1..2));
end proc;
```

Defines the operator multiplication by t.

```
L:=proc(f)
   t*f;
end proc;
```

Computes the first six approximations to the problem of Example 3.1 using CG and stores them in the sequence `Approx`.

```
Approx:=CG(L,sin(t),6,0.0000000001):
```

Plots each of the above approximations as dotted curves together with the true solution $sin(t)/t$ as a solid curve. It also outputs the i^{th} approximation `Approx[i]`, the i-th error `E[i]`, as well as the L-norm of `E[i]` in each case.

```
for i from 1 to 6 do
   plot([Approx[i],sin(t)/t],t=1..2,linestyle=[DOT,SOLID]):
   Approx[i];
   E[i]:=Approx[i]-sin(t)/t;
   LNormE[i]:=sqrt(InnProd(L(E[i]),E[i]));
end do;
```

Code for Example 3.2:

Defining the usual inner product for real-valued functions on $[0, \frac{\pi}{2}]$.

```
InnProd:=proc(f,g)
   evalf(int(f*g,t=0..Pi/2));
end proc;
```

Defines the operator corresponding to Example 3.2. For integration purposes we use a variable `dummyf` which is exactly the same as `f` except that it is an expression in terms of s instead of t.

```
L:=proc(f)
   local dummyf;
   dummyf:=subs(t=s,f);
   f-0.5*int(sin(s+t)*dummyf,s=0..Pi/2);
end proc;
```

Computes the first three approximations to the problem of Example 3.1 using CG and stores them in the sequence `Approx`.

```
Approx:=CG(L,1,3,0.0000000001):
```

Plots each of the above approximations as dotted curves together with the true solution in as a solid curve. It also outputs the i^{th} approximation `Approx[i]`, the i-th error `E[i]`, as well as the L-norm of `E[i]` in each case.

```
for i from 1 to 3 do
   plot([Approx[i],1+(sin(t)+cos(t))*4/(6-Pi)],t=0..Pi/2,linestyle=[DOT,SOLID]);
   Approx[i];
   E[i]:=evalf(Approx[i]-(1+(sin(t)+cos(t))*4/(6-Pi)));
   LnormOfE[i]:=sqrt(InnProd(L(E[i]),E[i]));
end do;
```

Code for Example 3.3:

Defining the usual inner product for real-valued functions on $[0, 1]$.

```
InnProd:=proc(f,g)
   evalf(int(f*g,t=0..1));
end proc;
```

Defines the operator corresponding to Example 3.3.

```
L:=proc(f)
   local dummyf;
   dummyf:=subs(t=s,f);
   f+5*int(s*dummyf,s=0..t)+5*t*int(dummyf,s=t..1);
end proc;
```

We choose the solution $u = e^t \cos(10t)$ and construct $f = Lu$.

```
solution:=(exp(t)*cos(10*t));
```

```
f:=simplify(L(solution));
```

We put f in numerical form before applying *CG*.

```
fnumeric:=evalf(f);
```

Computes the first six approximations to the problem of Example 3.3 using *CG* and stores them in the sequence `Approx`.

```
Approx:=CG(L,fnumeric,6,0.0000000001):
```

Plots the first six approximations as dotted curves together with the true solution as a solid curve. It also outputs the i^{th} approximation `Approx[i]`, the i-th error `E[i]`, as well as the L-norm of `E[i]` in each case.

```
for i from 1 to 6 do
   plot([Approx[i],solution],t=0..1,linestyle=[DOT,SOLID]):
   Approx[i];
   E[i]:=Approx[i]-solution;
   LnormOfE[i]:=sqrt(InnProd(L(E[i]),E[i]));
end do;
```

Since in the last two plots we cannot distinguish the approximation from the true curve, we plot the error `E[i]`.

```
plot(E[5],t=0..1);
plot(E[6],t=0..1);
```

Code for Example 3.4:

Defining the usual inner product for real-valued functions on $[-\infty, \infty]$. An important point is that the integration is done first symbolically and then evaluated numerically.

```
InnProd:=proc(f,g)
   evalf(int(f*g,t=-infinity..infinity));
end proc;
```

Defines the operator corresponding to Example 3.4.

```
L:=proc(f)
   local dummyf;
   dummyf:=subs(t=s,f);
   f+(1/sqrt(Pi))*int(exp(-(t-s)^2)*dummyf,s=-infinity..infinity);
end proc;
```

We choose the solution $u = e^{-t^2}\left(t^3 - \frac{3}{4}t\right)$ and construct $f = Lu$.

```
solution:=exp(-t^2)*(t^3-3/4*t);
```

```
f:=expand(L(solution));
```

We put f in numerical form before applying *CG*.

```
fnumeric:=evalf(f);
```

Computes the first six approximations to the problem of Example 3.4 using *CG* and stores them in the sequence `Approx`.

```
Approx:=CG(L,fnumeric,6,0.0000000001):
```

Plots the first six approximations as dotted curves together with the true solution as a solid curve. It also outputs the i^{th} approximation `Approx[i]`, the i-th error `E[i]`, as well as the L-norm of `E[i]` in each case.

```
for i from 1 to 6 do
   plot([Approx[i],solution],t=-6..6,linestyle=[DOT,SOLID]):
   Approx[i];
   E[i]:=Approx[i]-solution;
   LnormOfE[i]:=sqrt(InnProd(L(E[i]),E[i]));
end do;
```

Since in the last two plots we cannot distinguish the approximation from the true curve, we plot the error `E[i]`.

```
plot(E[5],t=-9..9);
plot(E[6],t=-9..9);
```

Code for Example 3.5:

Defining the usual inner product for real-valued functions on $[0, 1]$.

Input: f and g are expressions in terms of t.

Output: Numerical value of the inner product of f and g.

```
InnProd:=proc(f,g)
   evalf((int(f*g,t=1..2));
end proc;
```

Defines the operator multiplication by $\frac{1}{t}$.

```
L:=proc(f)
   f/t;
end proc;
```

Computes the first six approximations to the problem of Example 3.5 using CG and stores them in the sequence `Approx`.

```
Approx:=CG(L,t^2,6,0.0000000001):
```

Plots each of the above approximations as dotted curves together with the true solution t^3 as a solid curve. It also outputs the i^{th} approximation `Approx[i]`, the i-th error `E[i]`, as well as the L-norm of `E[i]` in each case.

```
for i from 1 to 6 do
   plot([Approx[i],t^3],t=1..2,linestyle=[DOT,SOLID]):
```

```
    Approx[i];
    E[i]:=Approx[i]-t^3;
    LNormE[i]:=sqrt(InnProd(L(E[i]),E[i]));
end do;
```

Bibliography

[1] N. I. Achieser. *Theory of Approximation.* Frederick Ungar Publishing Co., New York, 1956.

[2] T. M. Apostol. *Linear Algebra.* John Wiley and Sons Inc., New York, 1997.

[3] O. Axelsson. *Iterative Solution Methods.* Cambridge University Press, Cambridge, United Kingdom, 1994.

[4] O. Axelsson and I. Kaporin. On the sublinear and superlinear rate of convergence of conjugate gradient methods. *Numerical Algorithms*, 25(1), 2000.

[5] O. Axelsson and J. Karatson. On the rate of convergence of the conjugate gradient method for linear operators in hilbert space. *Numerical Functional Analysis and Optimization*, 23(4):285–302, 2002.

[6] P. N. Brown and H. F. Walker. Gmres on (nearly) singular systems. *SIAM Journal on Matrix Analysis and Applications*, 18(1):37–51, 1997.

[7] R. L. Burden and J. D. Faires. *Numerical Analysis.* Brooks Cole, seventh edition, 2000.

[8] S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer. Gmres and the minimal polynomial. *BIT*, 4:664–675, 1996.

[9] S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, C. D. Meyer, and Z. Q. Xue. Convergence estimates for solution of integral equations with gmres. *Journal of Integral Equations and Applications*, 8:19–34, 1996.

[10] J. W. Daniel. *The Approximate Minimization of Functionals.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.

[11] J. W. Demmel. *Applied Numerical Linear Algebra.* SIAM, Philadelphia, 1997.

[12] T. A. Driscoll, K.-C. Toh, and L. N. Trefethen. From potential theory to matrix iterations in six steps. *SIAM Review*, 40(3):547–578, 1998.

[13] G. H. Golub and C. F. V. Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore and London, third edition, 1996.

[14] A. Greenbaum. *Iterative Methods for Solving Linear Systems.* SIAM, Philadelphia, 1997.

[15] P. R. Halmos. *Introduction to Hilbert Space.* Chelsea Publishing Company, New York, second edition, 1972.

[16] P. R. Halmos and V. S. Sunder. *Bounded Integral Operators on L2 Spaces.* Springer-Verlag, New York, 1978.

[17] I. C. F. Ipsen and C. D. Meyer. The idea behind krylov methods. *The American Mathematical Monthly*, 105(10):889–899, 1998.

[18] L. V. Kantorovich and G. P. Akilov. *Functional Analysis.* Pergamon Press, Oxford, second edition, 1982.

[19] E. Kreyszig. *Introductory Functional Analysis with Applications.* John Wiley and Sons Inc., New York, 1978.

[20] S. Lipschutz and M. Lipson. *Linear Algebra.* McGraw-Hill, New York, third edition, 2001.

[21] O. Nevanlinna. *Convergence of Iterations for Linear Equations.* Birkhauser., Basel, Switzerland, 1993.

[22] W. K. Nicholson. *Linear Algebra with Applications.* PWS Publishing Company, Boston, 1995.

[23] D. Porter and D. S. Stirling. *Integral Equations: A Practical Treatment, from Spectral Theory to Applications.* Cambridge University Press, Cambridge, U.K., 1990.

[24] F. Riesz and B. Sz.-Nagy. *Functional Analysis.* Frederick Ungar Publishing Co., New York, second edition, 1965.

[25] Y. Saad and M.Schultz. Gmres: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat.Comp.*, 7(3):856–869, 1986.

[26] J. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. 1994.

[27] L. N. Trefethen and D. Bau. *Numerical Linear Algebra.* SIAM, Philadelphia, 1997.