Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Martin Tůma

From Moments to Modern Iterative Methods - Historical Connections and Inspirations

Katedra numerické matematiky

Vedoucí diplomové práce: prof. Ing. Zdeněk Strakoš, DrSc.
Studijní program: Matematika, Numerická a výpočtová matematika

2010

Na tomto místě bych rád poděkoval panu prof. Ing. Zdeňkovi Strakošovi, DrSc. za vedení diplomové práce a také za jeho podporu, trpělivost, rady, inspiraci a diskuze nejen při vypracování této diplomové práce. Rovněž patří můj dík rodičům a přítelkyni za podporu při studiu a tvorbu potřebného zázemí.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 2.8. 2010 Martin Tůma

Contents

Název práce: Od problému momentů k moderním iteračním metodám - historické souvislosti a inspirace
Autor: Martin Tůma
Katedra (ústav): Katedra numerické matematiky
Vedoucí bakalářské práce: prof. Ing. Zdeněk Strakoš, DrSc.
e-mail vedoucího: strakos@cs.cas.cz

Abstrakt: V této práci studujeme spojitosti mezi problémem momentů a moderními iteračními metodami. Uvedeme krátké shrnutí historie studia problému momentů. Ukážeme několik jeho definic a uvedeme motivace a výsledky několika významných matematiků, kteří se problémem momentů ve své práci zabývali. Dále ukážeme, jak spolu souvisí různé definice problém momentů, Gauss-Christoffelova kvadratura, teorie ortogonálních polynomů, řetězové zlomky, Sturm-Liouvillův problém, redukce modelu v lineárních dynamických systémech a některé iterační metody, jako je Lanczosova metoda a metoda sdružených gradientů.

Klíčová slova: momenty, Stieltjesův problém momentů, Vorobyevův problém momentů, Lanczosova metoda, metoda sdružených gradientů

Title: From Moments to Modern Iterative Methods - Historical Connections and Inspirations
Author: Martin Tůma
Department: Department of Numerical Mathematics
Supervisor: prof. Ing. Zdeněk Strakoš, DrSc.
Supervisor's e-mail address: strakos@cs.cas.cz

Abstract: In the present work we study the connections between the moment problem and the modern iterative methods. A short historical review of the study of the moment problem is given. Some different definitions of the moment problem are shown. Motivation and results of some mathematicians, who used the moment problem in their work are discussed. Connections between different definitions of the moment problem, Gauss-Christoffel quadrature, orthogonal polynomials, continued fractions, Sturm-Liouville problem, reduction of the model in linear dynamical systems and some of the iterative methods like Lanczos and Conjugate gradients method are explained.

Keywords: moments, Stieltjes moment problem, Vorobyev moment problem, Lanczos method, conjugate gradients method

Some basic notation

We will use linear operator A defined on the dense subset $\mathcal{D}(A)$ of the Hilbert space H with inner product $(.,.)$,

$$A : \mathcal{D}(A) \to H.$$

If not said otherwise we will use Hilbert spaces over $\mathbb{C}$, i.e.

$$(.,.) : H \times H \to \mathbb{C}.$$

The adjoint A^* of the operator A is defined in the following way: $y \in \mathcal{D}(A^*)$ if and only if we can find a $z \in H$ so that

$$(z, x) = (y, Ax), \quad \forall x \in \mathcal{D}(A).$$

Then we can set $A^*y = z$. In other words A^* is an adjoint of the operator A if and only if

$$(A^*y, x) = (y, Ax), \quad y \in \mathcal{D}(A^*), x \in \mathcal{D}(A).$$

The linear operator A is called bounded if and only if a positive number M exists such that

$$\|Ax\| \leq M\|x\|, \quad \forall x \in \mathcal{D}(A).$$

Graph $\mathcal{G}(A) \subset H \times H$ of the operator A is given by

$$\mathcal{G}(A) = \{[x, Ax] | x \in \mathcal{D}(A)\}.$$

Given operators A, S, we write $A \subset S$ and say S is an extension of A if and only if $\mathcal{G}(A) \subset \mathcal{G}(S)$.

An operator is called closed if and only if $\mathcal{G}(A)$ is a close subset of $H \times H$ and it is called closable if and only if $\overline{\mathcal{G}(A)}$ is the graph of an operator, in which case we define $\bar{A}$ the closure of A by $\mathcal{G}(\bar{A}) = \overline{\mathcal{G}(A)}$. So, $\mathcal{D}(\bar{A}) = \{z \in H | \exists x_n \in \mathcal{D}(A), x_n \to x, Ax_n$ is Cauchy $\}$ and $Ax = \lim_{n\to\infty} Ax_n$ can be set.

An operator A is called symmetric if $A \subset A^*$, self-adjoint if $A = A^*$ and essentially self-adjoint if $\bar{A}$ is self-adjoint.

The important case is, when H is of the finite dimension N. The linear operators on these finite dimensional spaces can be then represented by matrices $A \in \mathbb{C}^{N \times N}$. Matrix A is Hermitian if and only if $A = A^*$. Matrix A is positive definite if and only if $x^* A x > 0, \forall x \in H, x \neq 0$. We call matrix A HPD (Hermitian positive definite) if and only if A is Hermitian and positive definite.

Chapter 1

Introduction

This work is about the moment problem. During the last 150 years many books and papers have been published about this problem. It was P. Chebyshev who first started to study the moment problem. His motivations are described bellow this Chapter. Many mathematicians studied it from many different points of view. It is very interesting how many connections between the different parts of mathematics has been found in these works. As the time went on, the moment problem was used in order to solve various questions in mathematical statistics, theory of probability and mathematical analysis.

More about the history can be found in the Chapter 2 of this thesis. The purpose of the Chapter 2 is to briefly summarize the history and show how many different mathematicians touched the problem of moments in their works.

As the name of this work reminds, we will focus on the connection of the moment problem with the numerical linear algebra. The modern iterative methods can be seen as some kind of the model reduction using the moment problem.

In the Chapter 3 we will focus on one of the formulations of the moment problem, the simplified Stieltjes moment problem. The connection between this formulation of the moment problem, Gauss-Christoffel quadrature, orthogonal polynomials, continued fractions and the iterative methods like the Lanczos method [23], [24] and CG [19] will be described.

In the Chapter 4 we will review some results of Y. V. Vorobyev which he published in his book *Methods of Moments in Applied Mathematics* [41]. We will show his motivations and applications of his results. His ideas are now applied by many mathematicians in many different areas of mathematics, though the original book [41] remained almost unknown. This is why this Chapter is fairly extensive. Our intention is not only to show the connection of his work with the modern iterative methods but also give a short review of his contribution to other areas.

Both Chapters 3 and 4 have the similar aim, to show the connection between the different formulations of the moment problem with the modern iterative methods. Though the motivations of Stieltjes and Vorobyev were different, in the view of the modern iterative methods both formulations leads to the same result. And this result is, that the modern iterative methods based on the projections onto Krylov subspaces can be viewed as the model reduction which matches first n moments, where n depends on the iterative method.

In the Chapter 5 we will focus on the connection between the moment problem and the Sturm-Liouville problem. Although the connection is known, only a few works have been published about this topic. Our intention is to give a brief summary what is known about it and to contribute to this field by this little summarize review with the intention to underline the connection with the moment problem.

In the Chapter 6 we will look on the connection between the model reduction in linear dynamical systems and the moment problem. Some methods used in this area rediscovered within the last decades can be identified with the results published by Chebyshev and Stieltjes. Our intention is to summarize what is known about this topic.

Chapter 7 will be devoted to the numerical illustrations revealing the sensitivity of Gauss-Christoffel quadrature. In the Chapter 8 a brief conclusion of this work will be given.

Chapter 2

Problem of moments - historical background

Acording to J. A. Shoat and J. D. Tamarkin [33, p. 9] P. Chebyshev was the first who systematically discussed moment problem in his work [7]. For a detailed survey of Chebyshev life and work see e.g., [6]. In a series of papers started in 1855 he proposed the formulation of the following problem. Find a function $f(\lambda)$ such that

$$\int_a^b f(\lambda)\lambda^k d\lambda = \xi_k, \quad k = 0, 1, 2, ..., \tag{2.1}$$

where $\{\xi_k\}$ is a given sequence of numbers. Chebyshev was interested in the following two questions.

1) How far does a given sequence of moments determine the function $f(\lambda)$? More particularly, given

$$\int_a^b f(\lambda)\lambda^k d\lambda = \int_a^b e^{-\lambda^2}\lambda^k d\lambda, \quad k = 0, 1, 2, ...,$$

can we conclude that $f(\lambda) = e^{-\lambda^2}$? It would mean, that the distribution characterized by the function $\int_a^\lambda f(t)dt$ is a normal one.

2) What are the properties of the polynomials $\omega_n(z)$, denominators of successive

approximants to the continued fraction

$$\cfrac{a_0^2}{z-b_0-\cfrac{a_1^2}{z-b_1-\cfrac{a_2^2}{z-b_2-\cfrac{a_3^2}{z-b_3-\cfrac{a_4^2}{\ddots}}}}} \qquad (2.2)$$

The continued fractions (2.2) are nowadays known as J-fractions. This opened a vast new field, the general theory of orthogonal polynomials, of which only the classical polynomials of Legendre, Jacobi, Abel-Laguerre and Laplace-Hermite were known before Chebyshev. In his work we find numerous applications of orthogonal polynomials to interpolation, approximate quadratures and expansion of functions in series. Later they have been applied to the general theory of polynomials, theory of best linear approximation, theory of probability and mathematical statistics. For a more detailed view on the history of orthogonal polynomials with the intention to the connection with continued fractions see e.g., [5, pp. 213-224]. For a more recent overview of theory of orthogonal polynomials one can look at [40].

One of the Chebyshev's students A. Markov continued the work of his teacher. He was mainly interested in the theory of probability and he applied method of moments to the proof of the fundamental limit theorem. In 1884 in his thesis [27, pp. 172-180] Markov supplied a proof of the following so-called Chebyshev inequalities

$$\frac{\varphi(z_{l+1})}{\psi'(z_{l+1})}+...+\frac{\varphi(z_{n-1})}{\psi'(z_{n-1})}\leq\int_{z_l}^{z_n} f(\lambda)d\lambda\leq\frac{\varphi(z_l)}{\psi'(z_l)}+...+\frac{\varphi(z_n)}{\psi'(z_n)},$$

where $\varphi(z)/\psi(z)$ is one of the convergents of the continued fraction (2.2), and

$$z_1<z_2<...<z_l<z_{l-1}<...<z_{n-1}<z_n<...<z_m$$

are the roots of the equation $\psi(z)=0$. These inequalities were first given without proof by Chebyshev in 1874 in his work [8, pp. 157-160].

In 1896 Markov in his work [28, pp. 81-88] further generalized the moment problem by requiring the solution $f(\lambda)$ to be bounded

$$\int_{-\infty}^{\infty} f(\lambda)\lambda^k d\lambda=\xi_k, \quad k=0,1,2,..., \quad 0\leq f(\lambda)\leq L. \qquad (2.3)$$

Another research about moments was introduced by H. E. Heine in 1861 [18]. His work was motivated by the connection with continued fraction associated with

$$\int_a^b \frac{f(y)dy}{x-y}, \tag{2.4}$$

where the given function $f(y)$ is non-negative in (a,b) and also by an application of the orthogonal polynomials to approximate quadratures.

In the Chapter 3 of this thesis the Stieltjes's formulation of the problem of moments will be used. In 1894 T. J. Stieltjes published a classical paper *Recherches sur les fractions continues* [35]. He proposed and solved completely the following problem which he called the problem of moments. Find a bounded non-decreasing function $\omega(\lambda)$ in the interval $[0,\infty)$ such that its "moments"

$$\int_0^\infty \lambda^k d\omega(\lambda), \quad k=0,1,2,...,$$

have a prescribed set of values

$$\int_0^\infty \lambda^k d\omega(\lambda) = \xi_k, \quad k=0,1,2,.... \tag{2.5}$$

In this formulation of the moment problem Stieltjes used his own concept of integral - "Riemann-Stieltjes integral". The terminology was taken from mechanics. Consider the distribution function $\omega(\lambda)$ so that $\int_0^x d\omega(\lambda)$ represents the mass distributed over the segment [0,x], then the integrals

$$\int_0^\infty \lambda d\omega(\lambda), \int_0^\infty \lambda^2 d\omega(\lambda)$$

represent the first (statical) moment and the second moment (moment of inertia) with respect to 0 of the total mass $\int_0^\infty d\omega(\lambda)$ distributed over the real semi-axis $[0,\infty)$.

Stieltjes showed the necessary and sufficient condition for the existence of a solution of the problem of moments. Consider the Hankel matrices Δ_k and $\Delta_k^{(1)}$ of moments (2.5)

$$\Delta_k = \begin{bmatrix} \xi_0 & \xi_1 & \dots & \xi_k \\ \xi_1 & \xi_2 & \dots & \xi_{k+1} \\ \dots & \dots & \dots & \dots \\ \xi_k & \xi_{k+1} & \dots & \xi_{2k} \end{bmatrix}, \quad k=0,1,2..., \tag{2.6}$$

$$\Delta_k^{(1)} = \begin{bmatrix} \xi_1 & \xi_2 & \dots & \xi_k & \xi_{k+1} \\ \xi_2 & \xi_3 & \dots & \xi_{k+1} & \xi_{k+2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \xi_{k+1} & \xi_{k+2} & \dots & \xi_{2k} & \xi_{2k+1} \end{bmatrix}, \quad k = 0, 1, 2..., \tag{2.7}$$

then the solution of the Stieltjes moment problem exists if and only if the determinants $|\Delta_k|$ and $|\Delta_k^{(1)}|$ are positive. These determinants are known as Hankel determinants. In the more recent literature the proof can be found e.g. in [33, pp. 13-15]). The solution may be unique, in which case we speak of a determinate moment problem, or there may be more than one solution, in which case we speak of a indeterminate moment problem. Stieltjes made the solution of the moment problem (2.5) dependent upon the nature of the continued fraction corresponding to the integral

$$I(z, \omega) = \int_0^\infty \frac{d\omega(y)}{z - y}, \tag{2.8}$$

which corresponds to the following continued fraction

$$\cfrac{1}{c_1 z + \cfrac{1}{c_2 + \cfrac{1}{c_3 z + \cfrac{1}{c_4 + \cfrac{1}{\ddots}}}}} \tag{2.9}$$

Stieltjes showed that the moment problem (2.5) is determinate or indeterminate according as this continued fraction is convergent or divergent, that is, according as the series $\sum_{i=1}^{\infty} c_i$ diverges or converges. The continued fraction (2.9) is nowadays known as a Stieltjes fraction or S-fraction. It can be shown that this S-fraction can be transformed by contraction to a J-fraction (2.2) (see e.g., [3, pp. 2-3]) with $a_0^2 = 1/c_1, b_0 = -1/(c_1 c_2)$ and

$$a_k^2 = \frac{1}{c_{2k-1} c_{2k}^2 c_{2k+1}}, \quad b_k = -\frac{1}{c_{2k} c_{2k+1}} - \frac{1}{c_{2k+1} c_{2k+2}}, \quad k = 1, 2, ...$$

It took more than 20 years than another work about moment problem was introduced. It revived again in the work of H. Hamburger, R. Nevalina, M. Riesz, T. Carleman, F. Hausdorff and others, see e.g., [33] or the classical reference [1] for more detailed view on the history of the moment problem.

In the Chapter 5 of our thesis we will use the following extension of the work of Stieltjes to the whole real axis $(-\infty, \infty)$ which was introduced by H. Hamburger in 1920, see e.g., [16]. The problem is to determine a measure $\mu(x)$ such that

$$\int_{-\infty}^{\infty} x^k d\mu(x) = \xi_k, \quad k = 0, 1, 2.... \tag{2.10}$$

Again, if such a measure exists and is unique, the moment problem is determinate. If a measure μ exists but is not unique, the moment problem is called indeterminate. Hamburger showed that a necessary and sufficient condition for the existence of a solution of the Hamburger moment problem is the positivity of the determinants of the matrices (2.6), i.e., the positivity of the Hankel determinants. Proof can be found for example in [33, pp. 13-15].

In 1965 another approach was introduced by Russian mathematician Y. V. Vorobyev. In his book *Method of Moments in Applied mathematics* [41] he presented general problem of moments in Hilbert space. He applied his work about moments on solving differential, integral and finite difference equations and also on resolving spectrum of bounded operators in Hilbert space. As pointed out in Vorobyev [41, p. 113], his moment problem is very closely related with the method of determining the spectrum of A.N. Krylov, see e.g., [21] and [22]. In the Chapter 4 of this thesis it will be shown, that the only distinction between both methods is their computational scheme.

Chapter 3

Simplified Stieltjes moment problem

In this Chapter the formulation of a simplified Stieltjes moment problem will be given. This formulation will be used to show the connection between the moment problem, Gauss-Christoffel quadrature [9], Lanczos method [23], [24] and CG method [19]. Some results of theory of orthogonal polynomials and continued fractions which have very close connection to the moments will be shown. The connection between the above mathematical methods and the moment problem is known for a long time. The moments can be seen as the theoretical background for these methods. It is always useful to see things from many points of view and the moments can offer new insight. The elegant description of the connection which we will present in this thesis was given in the paper [38] by Z. Strakoš in 2009 and in the upcoming book [26] by J. Liesen and Z. Strakoš. We will start with brief introduction of some mathematical methods.

Gauss-Christoffel quadrature is one of the methods used for numerical integration of functions. Let $\omega(\lambda)$ be a non-decreasing distribution function on a finite interval $[a, b]$ on the real line. By the n-point Gauss-Christoffel quadrature we mean the approximation of the given Riemann-Stieltjes integral

$$I_\omega(f) = \int_a^b f(\lambda) d\omega(\lambda),$$

by the quadrature formula

$$I_\omega^n(f) = \sum_{l=1}^{n} \omega_l^{(n)} f(\lambda_l^{(n)}), \tag{3.1}$$

determined by the nodes $a \le \lambda_1^{(n)} < \lambda_2^{(n)}, ..., < \lambda_n^{(n)} \le b$ and positive weights

$$\omega_1^{(n)}, \omega_2^{(n)}, ..., \omega_n^{(n)},$$

such that $I_\omega^n(f) = I_\omega(f)$ whenever f is polynomial of a degree at most $2n-1$. In the following we will show, that it is the maximal number that can be reached.

Let the polynomial $\phi(\lambda) = \prod_{l=1}^n (\lambda - \lambda_l^{(n)})^2$. This polynomial ϕ is of a degree $2n$ and non-negative on the $[a, b]$. Obviously

$$\int_a^b \phi(\lambda) d\omega(\lambda) > 0,$$

but quadrature formula (3.1) gives

$$\sum_{l=1}^n \omega_l^{(n)} \phi(\lambda_l^{(n)}) = 0.$$

For the polynomial ϕ the quadrature formula (3.1) is not exact and so it could be exact for a polynomial of a degree at max $2n-1$.

The Lanczos method was introduced in 1950 by C. Lanczos, see [23] and [24]. Its symmetric variant is one of the most frequently used tools for computing a few dominant eigenvalues of a large sparse symmetric matrix A. It constructs a basis of Krylov subspaces which are defined for a square matrix A and a vector v by

$$\mathcal{K}_n(v, A) = \mathrm{span}(v, Av, ..., A^{n-1}v), \qquad n = 1, 2... \tag{3.2}$$

Starting from $v_1 = v, ||v|| = 1, v_0 = 0, \delta_1 = 0$

for $n = 1, 2, ...$

$$\begin{aligned} u_n &= Av_n - \delta_n v_{n-1}, \\ \gamma_n &= (u_n, v_n), \\ \hat{v}_{n+1} &= u_n - \gamma_n v_n, \\ \delta_{n+1} &= ||\hat{v}_{n+1}||, \quad \text{if} \quad \delta_{n+1} = 0 \quad \text{then stop}, \\ v_{n+1} &= \frac{\hat{v}_{n+1}}{\delta_{n+1}}. \end{aligned}$$

Since the natural basis $v, Av, ..., A^{n-1}v$ is ill-conditioned, the algorithm constructs an orthonormal basis of $\mathcal{K}_n(v, A)$. In matrix notation the Lanczos algorithm can be expressed as follows

$$AV_n = V_n T_n + \delta_{n+1} v_{n+1} (e^n)^T, \tag{3.3}$$

where

$$T_n = \begin{bmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_n \\ & & \delta_n & \gamma_n \end{bmatrix}, \quad \delta_l > 0, l = 2, ...n. \tag{3.4}$$

In 1952 a method of Conjugate gradients for solving linear systems was introduced by M. R. Hestenes and E. Stiefel in their joint paper [19]. Their intention was to provide an iterative algorithm for solving a system $Ax = b$. Generally the solution is given in N steps, but for matrices with some favorable properties it can give a good approximation to the solution in a few iterations. In the original paper Hestenes and Stiefel showed [19, pp. 21-25] the connection between the CG algorithm, Gauss quadrature, orthogonal polynomials and continued fractions.

Consider a HPD matrix A, given $x_0, r_0 = b - Ax_0, p_0 = r_0$, the subsequent approximate solutions x_n and the corresponding residual vectors $r_n = b - Ax_n$ are computed by

for $n = 1, 2, ...$

$$\begin{aligned} \gamma_{n-1} &= \frac{||r_{n-1}|||^2}{(p_{n-1}, Ap_{n-1})}, \\ x_n &= x_{n-1} + \gamma_{n-1}p_{n-1}, \\ r_n &= r_{n-1} - \gamma_{n-1}Ap_{n-1}, \\ \beta_n &= \frac{||r_n||^2}{||r_{n-1}||^2}, \\ p_n &= r_n + \beta_n p_{n-1}. \end{aligned}$$

Using the orthonormal basis V_n of the Krylov subspace $\mathcal{K}_n(A, v_1)$ determined by the Lanczos algorithm, one can write the approximation generated by the CG method in the following form

$$x_n = x_0 + V_n y_n.$$

We choose the next approximation x_n in order to r_n be orthogonal to V_n, i.e.,

$$0 = V_n^* r_n = V_n^*(b - Ax_n) = V_n^*(b - Ax_0 - AV_n y_n),$$

which gives, using $V_n^* A V_n = T_n$,

$$T_n y_n = ||r_0|| e_1$$

So, with the background of the Lanczos algorithm, the whole method can be formulated as

$$T_n y_n = ||r_0|| e_1, \quad x_n = x_0 + V_n y_n. \tag{3.5}$$

For a more detailed view on connections of Lanczos and CG and a description about the behavior of these algorithms in finite precision arithmetics see e.g., [29].

Now lets describe the simplified Stieltjes moment problem which will be used to show the connections between the methods presented above. Let $\omega(\lambda)$ be a non-decreasing distribution function defined on the interval $[a, b], \omega(a) = 0, \omega(b) = 1$. Given its moments

$$\xi_k = \int_a^b \lambda^k d\omega(\lambda), \quad k = 0, 1, ..., \tag{3.6}$$

we consider a problem to find a non-decreasing distribution function $\omega^{(n)}(\lambda)$ with n points of increase

$$\lambda_1^{(n)} < \lambda_2^{(n)} < ... < \lambda_n^{(n)},$$

and the associated positive weights $\omega_1^{(n)}, \omega_2^{(n)}, ..., \omega_n^{(n)}$, where $\sum_{l=1}^n \omega_l^{(n)} \equiv 1$, i.e.,

$$\omega^{(n)}(\omega) = \begin{cases} 0, & \text{if} \quad \lambda < \lambda_1^{(n)}, \\ \sum_{l=1}^i \omega_l^{(n)}, & \text{if} \quad \lambda_i^{(n)} \leq \lambda < \lambda_{i+1}^{(n)}, \\ \sum_{l=1}^n \omega_l^{(n)} \equiv 1, & \text{if} \quad \lambda_n^{(n)} \leq \lambda, \end{cases}$$

such that its first $2n$ moments match the first $2n$ moments (3.6) given by the distribution function $\omega(\lambda)$, i.e.,

$$\xi_k = \int_a^b \lambda^k d\omega(\lambda) = \sum_{l=1}^n \omega_l^{(n)} \{\lambda_l^{(n)}\}^k, \quad k = 0, 1, ..., 2n - 1. \tag{3.7}$$

Any polynomial is the linear combination of monomials λ^k. Therefore (3.7) means that the Riemann-Stieltjes integral with the distribution function $\omega(\lambda)$ of any polynomial up to degree $2n - 1$ is given by the weighted sum of the polynomial values at the n points $\lambda_l^{(n)}$ with the corresponding weights $\omega_l^{(n)}$. This is nothing but the n-point Gauss-Christoffel quadrature.

In [26, pp. 47-56] one can find an elegant description of some well known facts about orthogonal polynomials and continued fractions. Let

$$\psi_0(\lambda) = 1, \psi_1(\lambda), ..., \psi_n(\lambda)$$

be the first $n + 1$ orthogonal polynomials corresponding to the inner product

$$(\varphi, \psi) \equiv \int_a^b \varphi(\lambda) \psi(\lambda) d\omega(\lambda) \tag{3.8}$$

determined by the non-decreasing distribution function $\omega(\lambda)$, $\omega(a) = 0$, $\omega(b) = 1$, with the associated norm

$$\|\psi\|^2 \equiv (\psi, \psi) = \int_a^b \psi^2(\lambda) d\omega(\lambda).$$

And let

$$\varphi_0(\lambda) \equiv 1, \varphi_1(\lambda), ..., \varphi_n(\lambda)$$

be the first $n+1$ normalized orthonormal polynomials corresponding to the same inner product. These orthonormal polynomials satisfy the three-term Stieltjes recurrence

$$\delta_{n+1}\varphi_n(\lambda) = (\lambda - \gamma_n)\varphi_{n-1}(\lambda) - \delta_n \varphi_{n-2}(\lambda), \quad n = 1, 2, ..., \tag{3.9}$$

where $\delta_1 \equiv 0, \varphi_{-1} \equiv 0, \varphi_0(\lambda) \equiv 1$,

$$\gamma_n = \int_a^b (\lambda\varphi_{n-1}(\lambda) - \delta_n\varphi_{n-2}(\lambda))\varphi_{n-1}(\lambda) d\omega(\lambda)$$

and

$$\delta_{n+1} = \|(\lambda\varphi_{n-1}(\lambda) - \delta_n\varphi_{n-2}(\lambda)) - \gamma_n\varphi_{n-1}(\lambda)\|.$$

The above equations are given by the Gram-Schmidt orthogonalization process to the recursively generated sequence of polynomials $\lambda\varphi_0(\lambda) = \lambda, \lambda\varphi_1(\lambda), \lambda\varphi_2(\lambda)$ with the inner product (3.8). The details can be found in [26, p. 48]. Denoting by $\Phi_n(\lambda)$ a column vector with the orthonormal polynomials $\varphi_0(\lambda) \equiv 1, \varphi_1(\lambda), ..., \varphi_{n-1}(\lambda)$ as its entries

$$\Phi_n(\lambda) = [\varphi_0(\lambda) \equiv 1, \varphi_1(\lambda), ..., \varphi_{n-1}(\lambda)]^T,$$

the Stieltjes recurrence has the following matrix form

$$\lambda\Phi_n(\lambda) = T_n\Phi_n(\lambda) + \delta_{n+1}\varphi_n e_n.$$

The recurrence coefficients form a real symmetric tridiagonal matrix with positive subdiagonal, see (3.4). Now consider the continued fraction with the nth convergent given by

$$\mathcal{F}_n(\lambda) \equiv \cfrac{1}{\lambda - \gamma_1 - \cfrac{\delta_2^2}{\lambda - \gamma_2 - \cfrac{\delta_3^2}{\lambda - \gamma_{n-2} - \cfrac{\ddots}{\lambda - \gamma_{n-1} - \cfrac{\delta_n^2}{\lambda - \gamma_n}}}}} \equiv \frac{\mathcal{R}_n(\lambda)}{\mathcal{P}_n(\lambda)}.$$

If $\omega(\lambda)$ has a finite number of points of increase N, then the convergents form the finite sequence $\mathcal{F}_1(\lambda), \mathcal{F}_2(\lambda), ..., \mathcal{F}_N(\lambda)$, otherwise the sequence of convergents is infinite. The numerator and denominator of $\mathcal{F}_n(\lambda)$ are polynomials of degree $n-1$ and n.

The three-term recurrence (3.9) was originally published by Chebyshev in 1855, see [7]. His proof that the denominators of the convergents $\mathcal{F}_n(\lambda)$ of a continued fraction form a sequence of monic orthogonal polynomials associated with the distribution function $\omega(\lambda)$ may be regarded as the origin of a general theory of orthogonal polynomials, see Chapter 2 for more details of Chebyshev's work. In [26, pp. 52-53] the proofs of the following theorems about properties of the polynomials $\mathcal{F}_n(\lambda)$ are given.

Theorem 3.0.1 *Using the previous notation, for $n = 1, 2, ...$*

$$\mathcal{P}_n(\lambda) = \psi_n(\lambda),$$

$$\mathcal{R}_n(\lambda) = \int_a^b \frac{\psi_n(\lambda) - \psi_n(z)}{\lambda - z} d\omega(z) = \int_a^b \sum_{l=1}^{n} \frac{\psi_n^{(l)}(z)}{l!} (\lambda - z)^{l-1} d\omega(z).$$

Moreover, the numerators satisfy the three-term recurrence

$$\mathcal{R}_n(\lambda) = (\lambda - \gamma_n)\mathcal{R}_{n-1}(\lambda) - \delta_n^2 \mathcal{R}_{n-2}(\lambda), \quad n = 2, 3, ...$$

starting with $\mathcal{R}_0(\lambda) \equiv 0, \mathcal{R}_1(\lambda) \equiv 1$.

The following theorem describes the relationship between convergents of continued fractions, partial fractions and the Gauss-Christoffel quadrature.

Theorem 3.0.2 *Using the previous notation, the nth convergent $\mathcal{F}_n(\lambda)$ of the continued fraction corresponding to the non-decreasing distribution function $\omega(\lambda)$ can be decomposed into the partial fraction*

$$\mathcal{F}_n(\lambda) = \frac{\mathcal{R}_n(\lambda)}{\psi_n(\lambda)} = \sum_{j=1}^{n} \frac{\omega_j^{(n)}}{\lambda - \lambda_j^{(n)}}$$

where $\lambda_j^{(n)}$ and $\omega_j^{(n)}$ are the nodes and weights of the n-node Gauss-Christoffel quadrature associated with $\omega(\lambda)$, $j = 1, 2, ..., n$.

Consider the following decomposition for a sufficiently large λ

$$\frac{\omega_j^{(n)}}{\lambda - \lambda_j^{(n)}} = \frac{\omega_j^{(n)}}{\lambda} \left(1 - \frac{\lambda_j^{(n)}}{\lambda}\right)^{-1} = \sum_{l=1}^{2n} \omega_j^{(n)} \{\lambda_j^{(n)}\}^{l-1} \frac{1}{\lambda^l} + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}}\right),$$

and therefore

$$\mathcal{F}_n(\lambda) = \sum_{j=1}^{n} \frac{\omega_k^{(n)}}{\lambda - \lambda_j^{(n)}} = \sum_{l=1}^{2n} \frac{1}{\lambda^l} \left(\sum_{j=1}^{n} \omega_j^{(n)} \{\lambda_j^{(n)}\}^{l-1} \right) + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right).$$

Since $\omega_j^{(n)}$ and $\lambda_j^{(n)}$ represents the weights and nodes of the n-node Gauss-Christoffel quadrature associated with the distribution function $\omega(\lambda)$, the first $2n$ coefficients of the expansion of $\mathcal{F}_n(\lambda)$ into the power series around ∞ are equal to the first $2n$ moments,

$$\mathcal{F}_n(\lambda) = \sum_{l=1}^{2n} \frac{\xi_{l-1}}{\lambda^l} + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right),$$

$$\xi_{l-1} = \int_a^b \lambda^{l-1} d\omega(\lambda) = \sum_{j=1}^{n} \omega_j^{(n)} \{\lambda_j^{(n)}\}^{l-1}, \quad l = 1, 2, ...2n.$$

Chebyshev, Stieltjes and Heine considered expansion of the following integral (see the integrals (2.4) and (2.8) in the Chapter 2)

$$\int_a^b \frac{d\omega(u)}{\lambda - u}$$

into continued fractions. Using the similar expansions as above one can get

$$\frac{1}{\lambda - u} = \sum_{l=1}^{2n} \frac{u^{l-1}}{\lambda^l} + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right),$$

for sufficiently large λ the following formulas can be obtained

$$\int_a^b \frac{d\omega(u)}{\lambda - u} = \sum_{l=1}^{2n} \left(\int_a^b u^{l-1} d\omega(u) \right) \frac{1}{\lambda^l} + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right) = \sum_{l=1}^{2n} \frac{\xi_{l-1}}{\lambda^l} + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right)$$

i.e.,

$$\mathcal{F}_N(\lambda) = \int_a^b \frac{d\omega(u)}{\lambda - u} = \mathcal{F}_n(\lambda) + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right),$$

where $\mathcal{F}_n(\lambda)$ is the nth convergent of the continued fraction determined by the distribution function $\omega(\lambda)$

Now the formulation of the matching moment reduction in the matrix language will be given. Let's consider a linear algebraic system $Ax = b$ with a HPD matrix A

and an initial vector x_0, giving the initial residual $r_0 = b - Ax_0$ and the Lanczos initial vector $v \equiv v_1 = r_0/\|r_0\|$. Consider the non-decreasing distribution function $\omega(\lambda)$ with the points of increase equal to the eigenvalues of A and the weights equal to sizes of the squared components of v_1 in the corresponding invariant subspaces. For simplicity we assume that the eigenvalues of A are distinct and all weights are non-zero. With this setting, the moments (3.6) of the distribution function $\omega(\lambda)$ can be expressed in the matrix language as

$$\xi_k = \int_a^b \lambda^k d\omega(\lambda) = \sum_{l=1}^{N} \omega_l \{\lambda_l\}^k = v_1^* A^k v_1. \tag{3.10}$$

Let

$$\varphi_0(\lambda) \equiv 1, \varphi_1(\lambda), ..., \varphi_n(\lambda)$$

be the first $n+1$ orthonormal polynomials corresponding to the inner product

$$(\phi, \psi) = \int_a^b \phi(\lambda)\psi(\lambda) d\omega(\lambda)$$

determined by the distribution function $\omega(\lambda)$. Then again in a similar way as above by denoting

$$\Phi_n(\lambda) = [\varphi_0(\lambda), \varphi_1(\lambda), ..., \varphi_n(\lambda)]^T,$$

we get

$$\lambda \Phi_n(\lambda) = T_n \Phi_n(\lambda) + \delta_{n+1} \varphi_n(\lambda) e_n,$$

which represents the matrix formulation of the Stieltjes recurrence for the orthogonal polynomials. The recurrence coefficients form the Jacobi matrix T_n. Now consider the Gauss-Christoffel quadrature of

$$\int_a^b \lambda^k d\omega(\lambda).$$

The basic result about the Gauss-Christoffel quadrature (see e.g., [26, p. 61]) states that the nodes of the n-point quadrature are equal to the roots of $\varphi_n(\lambda)$, i.e., the eigenvalues of T_n. Using the spectral decomposition of the Jacobi matrix T_n and the well-known fact that all its eigenvectors have non-zero first components, we can see that the corresponding weights are given by the sizes of the squared first entries of the corresponding normalized eigenvectors of T_n.

$$\int_a^b \lambda^k d\omega^{(n)}(\lambda) = \sum_{l=1}^{n} \omega_l^{(n)} \{\lambda_l^{(n)}\}^k = e_1^T T_n^k e_1. \tag{3.11}$$

Lanczos algorithm applied to matrix A with v_1 (see (3.3)) gives in the nth step the same Jacobi matrix T_n which stores the coefficients of the Stieltjes reccurence (3.9), i.e.,

$$AV_n = V_n T_n + \delta_{n+1} v_{n+1} e_n^T$$

where V_n represents the matrix with orthonormal columns $v_1, v_2, .., v_n$,

$$V_n = [v_1, v_2, ..., v_n], \quad V_n^* V_n = I.$$

In [29, pp. 10-11] it is shown that the matrix T_n from the Lanczos process for A is the same as the matrix T_n in the Stieltjes reccurence for orthogonal polynomials corresponding to the inner product with the distribution function associated with A.

If we take a look at a matching moment model reduction at (3.7) which is equivalent to the Gauss-Christoffel quadrature and compare it with results in (3.10) and (3.11) we will arrive to the following summary. Let the original model is represented by the matrix A and the initial vector v_1. Then the Lanczos algorithm computes in steps 1 to n the model reduction of A with v_1 to T_n with e_1 such that the reduced model matches the first $2n$ moments of the original model, i.e.,

$$v_1^* A^k v_1 = e_1^* T_n^k e_1, \quad k = 0, 1, ..., 2n-1. \tag{3.12}$$

With A HPD, the considerations presented above can be extended to the CG method, see (3.5). The nth CG approximation x_n can be considered as a result of the model reduction from $Ax = b$ to $T_n y_n = \|r_0\| e_1$ such that the first $2n$ moments (3.12) are matched.

The similar results can be obtained also for methods for non-Hermitian matrices like non-Hermitian Lanczos method or Arnoldi method. In [38] and in the upcoming book [26] the Vorobyev moment problem is used to obtain these results. The Vorobyev moment problem and some of its applications will be the content of the following Chapter.

Chapter 4

Vorobyev moment problem

In this Chapter we will present the way how the moment problem was introduced and used in some applications by the Russian mathematician Y. V. Vorobyev in his book *The method of moments in applied mathematics* [41]. As stated on the webpage MathSciNet by American Mathematical Society, one can find only 7 mathematical works which has a citation of this book. This means, that the work of Vorobyev has been almost forgotten till today. The aim of this Chapter is to remind some of his ideas and present them in the way which reveal the connection with modern iterative methods.

4.1 Formulation of the moment problem and basic properties

Let $z_0, z_1, ..., z_n$ be $n+1$ prescribed linearly independent elements of Hilbert space H. Consider n-dimensional subspace H_n

$$H_n = \text{span}\{z_0, z_1, ..., z_{n-1}\}.$$

We wish to construct a linear operator A_n defined on the subspace H_n such that

$$\begin{aligned} A_n z_0 &= z_1, \\ A_n^2 z_0 &= z_2, \\ &\dots \\ A_n^{n-1} z_0 &= z_{n-1}, \\ A_n^n z_0 &= E_n z_n, \end{aligned} \tag{4.1}$$

where $E_n z_n$ is the projection of z_n on H_n. This problem is called the problem of moments in Hilbert space.
Let $x \in H_n$. By definition of H_n, it can be expressed as

$$x = c_0 z_0 + c_1 z_1 + ... + c_{n-1} z_{n-1},$$

where $c_0, c_1, ..., c_{n-1}$ are coefficients of linear combination. So

$$\begin{aligned} A_n x = c_0 A_n z_0 + c_1 A_n z_1 + ... + c_{n-1} A_n z_{n-1} = \\ = c_0 z_1 + c_1 z_2 + ... + c_{n-1} E_n z_n \in H_n. \end{aligned}$$

It means that A_n is completely defined by the system of equations (4.1). We will show that the eigenvalues of A_n are roots of polynomial P_n,

$$P_n(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + ... + \alpha_0, \tag{4.2}$$

where $\alpha_0, \alpha_1, ..., \alpha_{n-1}$ are such numbers that

$$E_n z_n = -\alpha_0 z_0 - \alpha_1 z_1 - ... - \alpha_{n-1} z_{n-1}. \tag{4.3}$$

Since $E_n z_n$ is an element of the subspace H_n, we can find such numbers $\alpha_0, \alpha_1, ..., \alpha_{n-1}$ that satisfy (4.3). This the use of (4.1) leads to

$$P_n(A_n) z_0 = (A_n^n + \alpha_{n-1} A_n^{n-1} + ... + \alpha_0 I) z_0,$$

with I the identity operator. Taking the scalar product of (4.3) successively with $z_0, z_1, ..., z_{n-1}$, we obtain a system of linear algebraic equations for the coefficients of the polynomial $P_n(\lambda)$.

$$\begin{gathered} (z_0, z_0)\alpha_0 + (z_0, z_1)\alpha_1 + ... + (z_0, z_{n-1})\alpha_{n-1} + (z_0, z_n) = 0 \\ (z_1, z_0)\alpha_0 + (z_1, z_1)\alpha_1 + ... + (z_1, z_{n-1})\alpha_{n-1} + (z_1, z_n) = 0 \\ ... \\ (z_{n-1}, z_0)\alpha_0 + (z_{n-1}, z_1)\alpha_1 + ... + (z_{n-1}, z_{n-1})\alpha_{n-1} + (z_{n-1}, z_n) = 0. \end{gathered} \tag{4.4}$$

In this we have made use of the fact that $z_n - E_n z_n$ must be orthogonal to every element of H_n, because $E_n z_n$ is the projection of z_n on H_n, i.e.,

$$(z_n - E_n z_n, z_k) = 0, \qquad k = 0, 1, ..., n-1,$$

so

$$(E_n z_n, z_k) = (z_n, z_k), \qquad k = 0, 1, ..., n-1.$$

The determinant of the system (4.4) is the Gramian of $z_0, z_1, ..., z_{n-1}$. It means that this system has a unique solution since their linear independence. Suppose now that λ is an eigenvalue and u an eigenelement of the operator A_n, that is

$$A_n u = \lambda u. \tag{4.5}$$

Since u belongs to the subspace H_n, it is expressible as

$$u = \xi_0 z_0 + \xi_1 z_1 + ... + \xi_{n-1} z_{n-1},$$

where $\xi_0, \xi_1, ..., \xi_{n-1}$ are scalars. Substituting this into equation (4.5) and making use of (4.1) and (4.3), we obtain

$$-\alpha_0 \xi_{n-1} z_0 + (\xi_0 - \alpha_1 \xi_{n-1}) z_1 + ... + (\xi_{n-2} - \alpha_{n-1} \xi_{n-1}) z_{n-1} =$$

$$= \lambda(\xi_0 z_0 + \xi_1 z_1 + ... + \xi_{n-1} z_{n-1}).$$

Since $z_0, z_1, ..., z_{n-1}$ are linearly independent, we can equate coefficients of like elements. This leads to the following system of equations

$$\begin{aligned} -\alpha_0 \xi_{n-1} &= \lambda \xi_0, \\ \xi_0 - \alpha_1 \xi_{n-1} &= \lambda \xi_1, \\ \xi_1 - \alpha_2 \xi_{n-1} &= \lambda \xi_2, \\ &... \\ \xi_{n-1} - \alpha_{n-1} \xi_{n-1} &= \lambda \xi_{n-1}, \end{aligned}$$

which can be written in the following form

$$\begin{bmatrix} -\lambda & & & & -\alpha_0 \\ 1 & -\lambda & & & -\alpha_1 \\ & 1 & -\lambda & & \vdots \\ & & \ddots & \ddots & \\ & & & 1 & -\alpha_{n-1} - \lambda \end{bmatrix} \begin{bmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \\ \xi_{n-1} \end{bmatrix} = 0.$$

In order for this system to have a non-trivial solution, its determinant must be equal to zero. Determinant of the above system can be written as $det(-\lambda I + C)$, where C is a companion matrix for $P_n(\lambda)$. So, it follows that eigenvalues of A_n are roots of $P_n(\lambda)$, i.e.,

$$P_n(\lambda) = \lambda^n + \alpha_{n-1} \lambda^{n-1} + ... + \alpha_0 = 0.$$

In [41, pp. 16-19] the way how to explicitly construct the inverse A_n^{-1} of the operator A_n is shown. It enables us to solve the following types of the equations

$$A_n x = f, \quad f \in H_n,$$

which is called the equation of the first kind, and

$$x = \mu A_n x + f, \quad f \in H_n,$$

where μ is the parameter and μ^{-1} is not an eigenvalue of the operator A_n. This equation is called the equation of the second kind.

Sequence of operators A_n may be constructed iteratively so as to converge to a preassigned bounded operator. Suppose that A is a bounded linear operator on the Hilbert space H, $\mathcal{D}(A) = H$ (if not said otherwise we will consider $\mathcal{D}(A) = H$ until the end of this Chapter). Choosing an element z_0, we form a sequence of the powers using the operator as follows

$$z_0, z_1 = Az_0, z_2 = Az_1 = A^2 z_0, ..., z_n = Az_{n-1} = A^n z_0.$$

By solving the moment problem(4.1) we determine a sequence of the operators A_n each defined on its own subspace H_n generated by all linear combinations of $z_0, z_1, ..., z_{n-1}$. The spaces expand with increasing n, and $H_n \subset H_{n+1}$. For each n the problem could be written in the following form

$$z_k = A^k z_0 = A_n^k z_0, \quad k = 0, 1, ..., n-1, \tag{4.6}$$

$$E_n z_n = E_n A^n z_0 = A_n^n z_0,$$

where $E_n A^n z_0$ is the projection of $A^n z_0$ on the subspace H_n. We will show that

$$A_n = E_n A E_n, \tag{4.7}$$

where E_n is the projection onto H_n. Any $x \in H_n$ can be written in the form

$$x = c_0 z_0 + c_1 z_1 + ... + c_{n-1} z_{n-1},$$

By definition, $E_n x = x$ and therefore

$$AE_n x = \sum_{k=0}^{n-1} c_k A z_k.$$

Applying (4.6) we obtain

$$AE_nx = \sum_{k=0}^{n-1} c_k A^{k+1} z_0 = \sum_{k=0}^{n-2} c_k A_n^{k+1} z_0 + c_{n-1} A^n z_0.$$

After applying E_n, we finally get

$$E_nAE_nx = \sum_{k=0}^{n-2} c_k A_n^{k+1} z_0 + c_{n-1} E_n A^n z_0 = \sum_{k=0}^{n-1} c_k A_n^{k+1} z_0 =$$

$$= \sum_{k=0}^{n-1} c_k A_n z_k = A_n x.$$

With this formula we can extend the domain of the operator A_n to the whole space H. We can also show that the sequence of the operators A_n is uniformly bounded

$$\|A_n\| = \|E_nAE_n\| \leq \|A\| \leq C.$$

In the following we will need the space H_z which we define as a closure of a linear manifold L_z consisting of elements of the form

$$x = Q(A)z_0,$$

where $Q(\lambda)$ is an arbitrary polynomial.

Theorem 4.1.1 *If A is a bounded linear operator and A_n a sequence of solutions of the moment problem (4.6), then the sequence A_n converges strongly to A in the subspace H_z.*

The proof can be found in [41, p. 21]. The fact that is possible to approximate bounded linear operators by operators like A_n means that they can be used to obtain solutions to various linear problems.

4.2 Application to the equations with completely continuous operators

An operator A is said to be degenerate if it can be represented in the form

$$Ax = \sum_{k=1}^{n} (x, \alpha_k)\beta_k,$$

with n finite and α_k, β_k given elements of the considered Hilbert space. A is then called completely continuous if for any positive number ϵ, it can be represented as

$$Ax = A_\epsilon^{'} + A_\epsilon^{''}x,$$

where $A_\epsilon^{'}$ is a degenerate operator and $\|A_\epsilon^{''}\| < \epsilon$. In the following we will show that a completely continuous operator is bounded. Let

$$A_\epsilon^{'}x = \sum_{k=1}^{n}(x, \alpha_k)\beta_k,$$

then

$$\|Ax\| \leq \sum_{k=1}^{n} |(x, \alpha_k)| \|\beta_k\| + \epsilon\|x\| \leq$$
$$\leq (\sum_{k=1}^{n} \|\alpha_k\|\|\beta_k\| + \epsilon)\|x\|,$$

so

$$\|A\| \leq \sum_{k=1}^{n} \|\alpha_k\|\|\beta_k\| + \epsilon.$$

Now we will show some examples of the completely continuous operators in Hilbert space. Because of the basic properties of the Hilbert spaces, in finite dimensional space H_n every linear operator is degenerate and therefore completely continuous. In infinite dimensional space l_2 any infinite matrix whose entries are such that

$$\sum_{i,k=1}^{\infty} |a_{i,k}| < \infty,$$

defines a completely continuous operator. In L_2 space, the integral operator

$$Ax = \int_a^b K(s,t)x(t)dt$$

is completely continuous if its kernel is square integrable, that is

$$\int_a^b \int_a^b |K(s,t)|^2 dsdt < \infty.$$

The next theorem is about the characterization of the completely continuous operators in the Hilbert space.

Theorem 4.2.1 *An operator A is completely continuous if and only if every infinite sequence of elements whose norms are uniformly bounded contains a subsequence $\{x_n\}$ for which the sequence $\{Ax_n\}$ is convergent.*

The theorem (4.1.1) can be strengthened for the completely continuous operators in the following way.

Theorem 4.2.2 *If A is a completely continuous operator, then the sequence of operators $\{A_n\}$ solving the moment problem (4.6) converges in norm to A in the subspace H_z.*

$$\lim_{n\to\infty} \|A - A_n\| = 0.$$

The proof can be found in [41, pp. 26-27]. Now consider the equation

$$x = \mu Ax + f, \tag{4.8}$$

where A is a completely continuous operator on H, $f \in H$ and μ is a parameter.

Theorem 4.2.3 *If $|\mu| \leq \|A\|^{-1}$, then the equation(4.8) has a solution for each given f. Such μ are then called the regular values.*

The proof can be found in [41, pp. 27-28]. In [41, pp. 27-37] theory of the moment problem is used to solve the equations of this type. We can set $z_0 = f$ and form the sequence of the powers

$$z_0 = f, z_1 = Af, ..., z_n = A^n f,$$

then we replace (4.8) by the approximate equation

$$x_n = \mu A_n x_n + f, \tag{4.9}$$

with A_n the solution of the problem of moments(4.6). The solution has the form

$$x_n = a_0 f + a_1 Af + ... + a_{n-1} A^{n-1} f,$$

where coefficients a_k are determined recursively

$$a_0 = 1 - \frac{\alpha_0}{P_n(\frac{1}{\mu})} \tag{4.10}$$

$$a_k = \mu a_{k-1} - \frac{\alpha_k}{P_n(\frac{1}{\mu})}, \qquad k = 1, 2, ..., n-1.$$

The quantities α_k are the coefficients of the polynomial $P_n(\lambda)$ (4.2) and satisfy the system (4.4). The following theorem is proved in [41, pp. 36-37]. It gives us an information how well can be solution x^* of (4.8) approximated by the solution x_n of (4.9).

Theorem 4.2.4 *If μ is a regular value for the equation*

$$x = \mu Ax + f \tag{4.11}$$

in which A is a completely continuous operator, then the sequence x_n of solutions of

$$x_n = \mu A_n x_n + f$$

converges to the solution x^* of equation (4.11) faster that a geometric progression with any arbitrarily small ration $q > 0$.

Now consider the equation

$$u - \mu Au = 0,$$

In [41, pp. 37-41] the formulation and proof of the following theorem in which the moment theory is used to determine its solution is given.

Theorem 4.2.5 *If A is a completely continuous operator and μ one of its reciprocal eigenvalues, that is, the equation*

$$u - \mu Au = 0$$

has a non-trivial solution in H_z, then each such solution is unique to within a factor, or in other words, to each eigenvalue there corresponds but one eigenelement in H_z. As n increases, one of the solutions of

$$u_n - \mu_n A_n u_n = 0$$

tends to the solution of $u - \mu Au = 0$ faster than a geometric progression with any arbitrary small ratio $q > 0$.

In [41, p. 44] it is shown that the eigenvalues of the operator A_n can be found as roots of polynomial $P_n(\frac{1}{\mu})$

This theory could be used on the following example as shown in [41, pp. 44-47]. A problem arising in heat conduction requires the solution of the equation

$$-y''(x) = \mu y(x)$$

under the boundary conditions

$$y(0) = 0, y(1) + y'(1) = 0.$$

By means of Green's function, this problem is easily converted into an integral equation

$$y = \mu \int_0^1 K(x,\xi) y(\xi) d\xi,$$

where

$$K(x,\xi) = \begin{cases} \frac{1}{2}(2-\xi)x, & x \le \xi, \\ \frac{1}{2}(2-x)\xi, & x \ge \xi. \end{cases}$$

The operator

$$Ay = \int_0^1 K(x,\xi) y(\xi) d\xi$$

is completely continuous in $L_2([0,1])$ space. For the function z_0 we choose

$$z_0 = \sin \pi x.$$

The sequence z_k is then constructed by successive integration

$$z_{k+1}(x) = \int_0^1 K(x,\xi) z_k(\xi) d\xi.$$

We compute z_1, z_2, z_3 and also scalar products (z_l, z_k) needed in the system (4.4). Then the $\alpha_0, \alpha_1, \alpha_2$ are computed and used for the equations $P_1(\frac{1}{\mu}) = 0, P_2(\frac{1}{\mu}) = 0, P_3(\frac{1}{\mu}) = 0$. Three different eigenvalues are computed and compared with the values from another numerical method and with the exact values. Details can be found in [41, p. 47].

4.3 Application to the equations with self-adjoint operators

Now let A be the bounded self-adjoint operator with $\mathcal{D}(A) = H$. The bounds of a self-adjoint operator are the smallest number m and largest number M for which

$$m(x,x) \le (Ax,x) \le M(x,x)$$

for every $x \in H$. If $m(x,x) > 0$ then we call the operator A positive definite. We first consider a self-adjoint operator A_n in some n-dimensional space H_n. An operator A_n is self-adjoint, so we can do its spectral decomposition

$$A_n U = \Lambda U,$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}, \quad U = [u_1, u_2, ..., u_n].$$

The eigenvectors $[u_1, u_2, ..., u_n]$ may be taken as orthonormal basis of H_n. So each element $x \in H_n$ can be represented as

$$x = (x, u_1)u_1 + (x, u_2)u_2 + ... + (x, u_n)u_n.$$

We now introduce the notion of spectral function of an operator used in Vorobyev's book (it is nothing else than different name for distribution function associated with operator A_n used in previous Chapters). Let us arrange the eigenvalues of A_n in increasing order of magnitude, i.e.,

$$\lambda_1 < \lambda_2 < ... < \lambda_n$$

The spectral function of A_n, denoted by δ_λ, is defined to be the family of projections determined by the set of relations

$$\begin{array}{ll} \delta_\lambda x = 0 & \lambda < \lambda_1, \\ \delta_\lambda x = (x, u_1)u_1 & \lambda_1 \leq \lambda < \lambda_2, \\ \delta_\lambda x = (x, u_1)u_1 + (x, u_2)u_2 & \lambda_2 \leq \lambda < \lambda_3, \\ \cdots & \cdots \\ \delta_\lambda x = (x, u_1)u_1 + ... + (x, u_{n-1})u_{n-1} & \lambda_{n-1} \leq \lambda < \lambda_n, \\ \delta_\lambda x = (x, u_1)u_1 + ... + (x, u_n)u_n = x & \lambda \geq \lambda_n, \end{array}$$

for any element $x \in H_n$. The real function $(\delta_\lambda x, x)$ has the form

$$\begin{array}{ll} (\delta_\lambda x, x) = 0 & \lambda < \lambda_1, \\ (\delta_\lambda x, x) = |(x, u_1)|^2 & \lambda_1 \leq \lambda < \lambda_2, \\ \cdots & \cdots \\ (\delta_\lambda x, x) = \sum_{k=1}^{n-1} |(x, u_k)|^2 & \lambda_{n-1} \leq \lambda < \lambda_n, \\ (\delta_\lambda x, x) = \|x\|^2 & \lambda \geq \lambda_n. \end{array}$$

So, $(\delta_\lambda x, x)$ is constant on those ranges of λ where A_n has no eigenvalues and increases by jumps at each eigenvalue λ_k equal to the amount $|(x, u_k)|^2$. If we let $\triangle_k \delta_\lambda$ denote the value of the jump in the projection,

$$\triangle_k \delta_\lambda x = (x, u_k)u_k,$$

then the expression

$$A_n x = \lambda_1(x, u_1)u_1 + \lambda_2(x, u_2)u_2 + ... + \lambda_n(x, u_n)u_n$$

can be written in the form

$$A_n x = \sum_{k=1}^{n} \lambda_k \triangle_k \delta_\lambda x.$$

A similar formula holds for any arbitrary bounded self-adjoint operator A, i.e.,

$$Ax = \int_m^M \lambda d\delta_\lambda x,$$

in which the spectral function δ_λ of the operator A is understood to be a family of projections depending on the real parameter λ such that

1) δ_λ is non-decreasing with increasing λ, that is, if $\mu > \lambda$, then the subspace into which δ_μ projects contains the subspace into which δ_λ projects,

2) $\delta_m = 0$, $\delta_M = I$,

3) δ_λ is continuous from right, i.e.,

$$\lim_{\lambda \to \lambda' +} \delta_\lambda = \delta_{\lambda'}.$$

The self-adjointness of a given operator A greatly simplifies the solution of the problem of moments and makes possible the development of useful algorithms for solving corresponding linear problems. As pointed out in [41, pp. 53-54] in the case of a self-adjoint operator, our problem of moments for operators is completely equivalent to the classical Chebyshev, Markov and Stieltjes scalar problem of moments (see the Chapter 2, especially (2.1), (2.3) and (2.5)) for the quadratic functional (Ax, x) since we can connect operator A to its corresponding quadratic functional (Ax, x)

$$(Ax, x) = \int_m^M \lambda d(\delta_\lambda x, x).$$

Function $d(\delta_\lambda x, x)$ here is nothing else than a continuous version of $d\omega(\lambda)$ used in (3.10) in the Chapter 3. As pointed out in [26] the above integral representation of self-adjoint operators has been used in mathematical foundation of quantum mechanics which has roots in works of J. Neumann, D. Hilbert and A. Wintner, see [26, pp. 79-80] for details and references of the original articles.

Let A be a prescribed self-adjoint bounded linear operator in Hilbert space H. Starting with an arbitrarily chosen element z_0, we form the sequence of iterations

$$z_0, z_1 = Az_0, ..., z_n = Az_{n-1} = A^n z_0, ...$$

The solution of the problem of moments (4.6) gives a sequence of operators A_n, each defined on its own subspace H_n generated by all possible linear combinations of $z_0, z_1, ..., z_{n-1}$. Since $A_n = E_n A E_n$, A_n is also self-adjoint. To determine its spectrum Lanczos method of successive orthogonalization is used

$$p_0 = z_0$$

$$p_{k+1} = (A - a_k I)p_k - b_{k-1}p_{k-1},$$

where

$$a_k = \frac{(Ap_k, p_k)}{(p_k, p_k)}, b_{k-1} = \frac{(p_k, Ap_{k-1})}{(p_{k-1}, p_{k-1})} = \frac{(p_k, p_k)}{(p_{k-1}, p_{k-1})}.$$

The elements p_k are mutually orthogonal, i.e., $(p_k, p_l) = 0$ for $k \neq l$.

$$p_k = P_k(A) z_0,$$

with $P_k(A)$ a kth degree polynomial in A. The polynomials $P_k(\lambda)$ have leading coefficients of one and satisfy the following recursion relations

$$P_{k+1}(\lambda) = (\lambda - a_k)P_k(\lambda) - b_{k-1}P_{k-1}(\lambda),$$

$$b_{-1} = 0, P_0 = 1.$$

This is nothing else than Stieltjes recurrence for orthogonal polynomials (3.9) mentioned in Chapter 3. We have already shown that roots of

$$P_n(\lambda) = 0$$

are the eigenvalues of the operator A_n. For the case of self-adjoint operators we have the following theorem to determine its eigenvalues

Theorem 4.3.1 *If A is a bounded self-adjoint operator and A_n a sequence of solutions of the moment problem (4.6), then the sequence of spectral functions $\delta_\lambda^{(n)}$ of the operators A_n converges strongly to the spectral function of A*

$$\delta_\lambda^{(n)} \to \delta_\lambda, \qquad n \to \infty,$$

in the subspace H_z for all λ not belonging to the discrete spectrum of A.

The proof can be found in [41, p.61]. For the case of the non-homogeneous equations with self-adjoint operator A Vorobyev proved the following theorem

Theorem 4.3.2 *Let A be a bounded self-adjoint positive definite operator defined on the Hilbert space H and let z_0 be any element of H. Then we can create a sequence x_n*

$$\begin{aligned} x_{n+1} &= x_n + h_n F_n(0) g_n, \\ g_n &= r_n + l_{n-1} g_{n-1}, \qquad h_{n-1} = \frac{\|r_{n-1}\|^2}{(Ag_{n-1}, g_{n-1})}, \\ r_n &= r_{n-1} - h_{n-1} A g_{n-1}, \qquad l_{n-1} = \frac{\|r_n\|^2}{\|r_{n-1}\|^2}, \\ F_n(0) &= F_{n-1}(0) + \frac{(f, r_n)}{\|r_n\|^2}, \\ r_0 &= g_0 = z_0, F_0(0) = \frac{(f, z_0)}{\|z_0\|^2}, \end{aligned}$$

which converges strongly to the solution x of equation

$$Ax = f,$$

where f is in the subspace H_z.

The algorithm in the above theorem is nothing else than the conjugate-gradients algorithm (see (3.5)) which was first published by M. Hestenes and E. Stiefel in their joint paper [19]. The proof of the theorem can be found in [41, 65-70].

4.4 Solution of time-dependent problems

Vorobyev applied his method on various problems. In this section we will briefly show, how it can be applied in solving various time-dependent problems. We begin with the solution of the Cauchy problem for a partial differential equation of first order

$$\frac{\partial x}{\partial t} = -Ax, \tag{4.12}$$

with $x \in H$ and A a positive definite symmetric linear operator defined on a dense subset $\mathcal{D}(A)$ of H. In general, A is an unbounded operator. The initial data for (4.12) is

$$x(0) = x_0 \in H.$$

Since A is positive definite, it possesses a bounded inverse, which we shall assume to be defined over all of H. We can then write (4.12) in the form

$$A^{-1}\frac{\partial x}{\partial t} + x = 0. \tag{4.13}$$

Setting $z_0 = x_0, z_1 = A^{-1}z_0, ..., z_n = A^{-1}z_{n-1}, ...,$ we construct the sequence A_n of solutions of the problems of moments (4.6), and we replace the equation (4.13) by the equation

$$A_n\frac{\partial x_n}{\partial t} + x_n = 0 \tag{4.14}$$

with the initial condition $x_n(0) = x_0$. If we introduce the spectral function δ_λ of the inverse operator A^{-1}, the required solution can be expressed as

$$x = \int_0^M e^{-t/\lambda} d\delta_\lambda x_0,$$

with M the least upper bound of the spectrum of A^{-1} and so equal to $\|A^{-1}\|$.

Since $e^{-t/\lambda}$ is continuous for $\lambda \geq 0$ for all $t \geq 0$ and since the strong convergence of a sequence of operators implies the strong convergence of a continuous function of them, it follows that the sequence x_n of solutions of (4.14) converges strongly to x, i.e.,

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

Let us next show how the approximate solution x_n may be computed. Set

$$x_n = \eta_0(t)z_0 + \eta_1(t)z_1 + ... + \eta_{n-1}z_{n-1},$$

and substitute this into equation (4.14). Using the formulas (4.1) we obtain

$$\eta_0(t)z_0 + (\eta_1 + \frac{d\eta_0}{dt})z_1 + ... + (\eta_{n-1} + \frac{d\eta_{n-2}}{dt}) + \frac{d\eta_{n-1}}{dt}E_n z_n = 0.$$

The application of the expression

$$E_n z_n = -\alpha_0 z_0 - \alpha_1 z_1 - ... - \alpha_{n-1}z_{n-1}$$

gives

$$(\eta_0 - \alpha_0\frac{d\eta_{n-1}}{dt})z_0 + (\eta_1 + \frac{d\eta_0}{dt} - \alpha_1\frac{d\eta_{n-1}}{dt})z_1 + ... + (\eta_{n-1} + \frac{d\eta_{n-2}}{dt} - \alpha_{n-1}\frac{d\eta_{n-1}}{dt})z_{n-1} = 0.$$

Since $z_0, z_1, ..., z_{n-1}$ are by assumption linearly independent, by equating their coefficients to zero, we arrive at a system of these scalar equations equivalent to (4.14)

$$\begin{aligned} \eta_0 - \alpha_0 \frac{d\eta_{n-1}}{dt} &= 0, \\ \frac{d\eta_0}{dt} + \eta_1 - \alpha_1 \frac{d\eta_{n-1}}{dt} &= 0, \\ &\cdots \\ \frac{d\eta_{n-2}}{dt} + \eta_{n-1} - \alpha_{n-1} \frac{d\eta_{n-1}}{dt} &= 0, \end{aligned} \tag{4.15}$$

with corresponding initial conditions

$$\eta_0(0) = 1, \eta_1(0) = \eta_2(0) = ... = \eta_{n-1}(0) = 0.$$

Vorobyev showed in [41, pp. 98-99] that equations (4.15) may be solved by means of Laplace transforms. Let

$$\xi_k = \int_0^\infty \eta_k e^{-\lambda t} dt.$$

Then

$$\begin{aligned} \xi_0 - \alpha_0 \lambda \xi_{n-1} &= 0, \\ \lambda \xi_0 + \xi_1 - \alpha_1 \lambda \xi_{n-1} &= 1, \\ &\cdots \\ \lambda \xi_{n-2} + \xi_{n-1} - \alpha_{n-1} \lambda \xi_{n-1} &= 0, \end{aligned}$$

so

$$\begin{aligned} \xi_j &= \frac{P_{n-j-1}(-\frac{1}{\lambda})}{\lambda^2 P_n(-\frac{1}{\lambda})}, \quad j \geq 1, \\ \xi_0 &= \frac{\alpha_0}{\lambda P_n(-\frac{1}{\lambda})}. \end{aligned}$$

Here $P_n(v)$,

$$P_n(v) = v^n + \alpha_{n-1} v^{n-1} + ... + \alpha_0$$

is the nth orthogonal polynomial, the roots of which are all real and lie in the interval $[0, M]$. The solution of (4.15) is then given by the contour integrals

$$\begin{aligned} \eta_j(t) &= \frac{1}{2\pi i} \lim_{T \to \infty} \int_{\sigma - iT}^{\sigma + iT} \frac{P_{n-j-1}(-\frac{1}{\lambda})}{\lambda^2 P_n(-\frac{1}{\lambda})} e^{\lambda t} d\lambda, \quad j \geq 1, \\ \eta_0(t) &= \frac{1}{2\pi i} \lim_{T \to \infty} \int_{\sigma - iT}^{\sigma + iT} \frac{\alpha_0}{\lambda P_n(-\frac{1}{\lambda})} e^{\lambda t} d\lambda. \end{aligned}$$

The above integrals may be evaluated by means of the residues.

The non-homogeneous problem

$$\frac{\partial x}{\partial t} = -Ax + gf(t), \tag{4.16}$$

$x(0) = 0$, with $g \in H$ and $f(t)$ a scalar function, may be solved in a similar way. The desired solution is

$$x = \int_0^M [\int_0^t f(\xi)e^{-(t-\xi)/\lambda}d\xi]d\delta_\lambda g$$

and it is a continuous function of the inverse operator. Therefore, the sequence x_n of solutions of

$$A_n\frac{\partial x_n}{\partial t} + x_n = A^{-1}gf(t), \tag{4.17}$$

with A_n the operators solving the problems of moments (4.6) defined by $z_0 = g, z_1 = A^{-1}g, ..., z_n = A^{-1}z_{n-1}, ...$, converges strongly to x when n approaches infinity. Again let us have a look for an approximate solution x_n of the form

$$x_n = \eta_0(t)z_0 + \eta_1(t)z_1 + ... + \eta_{n-1}z_{n-1}.$$

The substitution of this in equation (4.17) leads to a system of scalar equations

$$\begin{aligned} \eta_0 - \alpha_0\frac{d\eta_{n-1}}{dt} &= 0, \\ \frac{d\eta_0}{dt} + \eta_1 - \alpha_1\frac{d\eta_{n-1}}{dt} &= 0, \\ &... \\ \frac{d\eta_{n-2}}{dt} + \eta_{n-1} - \alpha_{n-1}\frac{d\eta_{n-1}}{dt} &= 0, \end{aligned}$$

which can be solved by Laplace transform to obtain

$$\begin{aligned} \eta_j(t) &= \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\sigma-iT}^{\sigma+iT}\frac{F(\lambda)P_{n-j-1}(-\frac{1}{\lambda})}{\lambda^2 P_n(-\frac{1}{\lambda})}e^{\lambda t}d\lambda, \quad j \geq 1, \\ \eta_0(t) &= \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\sigma-iT}^{\sigma+iT}\frac{\alpha_0 F(\lambda)}{\lambda P_n(-\frac{1}{\lambda})}e^{\lambda t}d\lambda, \end{aligned}$$

where

$$F(\lambda) = \int_0^\infty f(t)e^{-\lambda t}d\lambda.$$

By a very similar method we can find a numerical solution for oscillation problem characterized by a partial differential equation of second order with respect to time

$$\frac{\partial^2 x}{\partial t^2} = -Ax, \tag{4.18}$$

with A a positive symmetric operator as before and with initial data prescribed at $t = 0$

$$x(0) = x_0, \quad \frac{\partial x(0)}{\partial t} = x_0^{'}.$$

And also we can find a numerical solution for the non-homogeneous equation with homogeneous initial conditions.

$$\frac{\partial^2 x}{\partial t^2} = -Ax + gf(t),$$

with $g \in H$ and $f(t)$ a scalar function. Detailed description of solutions of the problems above can be found in [41, pp. 100-105].

This method can be also applied if the equation describing the small oscillations has the form

$$M\frac{\partial^2 x}{\partial t^2} + Ax = 0, \tag{4.19}$$

where M is positive definite. The equation (4.19) can be reduced to the equation (4.18). It can be written in the following form

$$\frac{\partial^2 x}{\partial t^2} + M^{-1}Ax = 0.$$

However, the operator M^{-1} is non-symmetric. By a change in the definition of a scalar product, $A^{'} = M^{-1}A$ can be turned into a symmetric operator. The new space $H^{'}$ can be introduced. It consists of the same elements as H but the scalar product in it is defined as follows

$$[x, y] = (x, My).$$

The parentheses denote the scalar product in the original space H, while the brackets denote it in the new space $H^{'}$. The symmetry of $A^{'}$ is then implied by the equalities

$$[A^{'}x, y] = (A^{'}x, My) = (Ax, y) = (x, Ay) = [x, M^{-1}Ay] = [x, A^{'}y].$$

Now it will be shown, how Vorobyev used the moment problem in order to reduce the order of the system of linear differential equations. Our intention is to replace the given system by an approximate one close to it but of lower order. Vorobyev in his book [41, p. 106] said that

"this problem at first appears to be meaningless. An Nth order system has N eigenoscillations each of which may be excited by the application of a suitable choice of external forces to the physical system. In the reduction of the order of the system, not only would the eigenoscillations be distorted but they would also be reduced in number, so some of them would disappear. However for real system, the external forces are far from being arbitrary. Usually one or several members of a physical system are subjected to perturbations, the other being unaffected and displaced only as a result of their relationship to the other members. Under these conditions, certain eigenoscillations may eighter not be excited or be excited so little as to have no appreciable effect on the overall oscillations. Under these circumstances, the problem of reducing the order of a system of differential equations is now a meaningful and a very urgent one since the order of a system is often determined not by the physical essence of the problem but by the degree of idealization that we have adopted."

This is an essential idea in all methods based on the reduction of the system. When we solve the linear system

$$Ax = b$$

by some modern iterative method like CG or GMRES we take into consideration that the matrix A and the vector b are not arbitrary, but they all come from a real world problem and have strong relationship to each other.

But lets go back to the system of linear differential equations. Let x be a vector in N-dimensional space H_N. We shall suppose that the system of equations has this normal form

$$\frac{dx}{dt} = Ax + gf(t). \tag{4.20}$$

The initial conditions are assumed to be zero. To solve (4.20), the method of moments can be applied. Set $z_0 = g$ and form the iterations

$$z_0 = g, z_1 = Az_0, ..., z_n = A^n z_0 = Az_{n-1}, ...$$

We suppose that $z_0, z_1, ..., z_n$ are linearly independent. We can again denote A_n the solutions of the moment problem (4.6) each defined on H_n. The eigenvalues of A_n are roots of $P_n(\lambda)$ (4.2). As shown earlier, $\lim_{n\to\infty} A_n = A$, so an approximate solution to (4.20) can be found by solving

$$\frac{dx_n}{dt} = A_n x_n + gf(t). \tag{4.21}$$

Now x_n belongs to H_n and can be represented as

$$x_n = \eta_0 z_0 + \eta_1 z_1 + ... + \eta_{n-1} z_{n-1}, \tag{4.22}$$

with $\eta_0(t), \eta_1(t), ..., \eta_{n-1}(t)$ scalar functions. The substitution of (4.22) into (4.21) and the application of formula (4.3) then leads to

$$(\frac{d\eta_0}{dt} + \alpha_0 \eta_{n-1}) z_0 + (\frac{d\eta_1}{dt} - \eta_0 + \alpha_1 \eta_{n-1}) z_1 + ... + (\frac{d\eta_{n-1}}{dt} - \eta_{n-2} + \alpha_{n-1}\eta_{n-1}) z_{n-1} = z_0 f(t).$$

If we equate the coefficients of like elements, we arrive at the following system of scalar equations

$$\frac{d\eta_0}{dt} + \alpha_0 \eta_{n-1} = f(t),$$
$$\frac{d\eta_1}{dt} - \eta_0 + \alpha_0 \eta_{n-1} = 0,$$
$$...$$
$$\frac{d\eta_{n-1}}{dt} - \eta_{n-2} + \alpha_{n-1} \eta_{n-1} = 0,$$

with initial conditions $\eta_0(0) = \eta_1(0) = ... = \eta_{n-1}(0) = 0$. This system is equivalent to the equation

$$\eta_{n-1}^{(n)} + \alpha_{n-1} \eta_{n-1}^{(n-1)} + ... + \alpha_0 \eta_{n-1} = f(t)$$

with the initial conditions $\eta_{n-1}(0) = \eta_{n-1}^{(1)}(0) = ... = \eta_{n-1}^{(n-1)}(0) = 0$. The differential equation above may be solved by any method for solving a ordinary differential equations.

Now let's look at the error of this approximation x_n, let y_n denote the error in the approximate solution, that is

$$x = x_n + y_n.$$

Substitution of this in (4.20) gives

$$\frac{d\eta_0}{dt} z_0 + \frac{d\eta_0}{dt} z_1 + ... + \frac{d\eta_{n-1}}{dt} z_{n-1} + \frac{dy_n}{dt} =$$

$$= \eta_0 z_1 + \eta_1 z_2 + ... + \eta_{n-1} E_n z_n + \eta_{n-1}(z_n - E_n z_n) + A y_n + g f(t),$$

which implies the following equation for the error

$$\frac{dy_n}{dt} = A y_n + \eta_{n-1}(t)(z_n - E_n z_n),$$

with initial conditions

$$y_n(0) = \frac{dy_n(0)}{dt} = ... = \frac{d^n y_n(0)}{dt^n} = 0.$$

The error is proportional to $\|z_n - E_n z_n\|$. If $\|z_n - E_n z_n\|$ is small the error will be also small. Two of the important cases in which the quantity $\|z_n - E_n z_n\|$ is small for $n < N$ and the system of differential equations (4.20) can be approximated by the lower order system (4.21) will be indicated. For simplicity, suppose that the operator A has a simple structure. Then the vector g, which represent the amplitude of the external force, can be expanded in terms of its eigenvectors

$$g = \sum_{k=1}^{N} a_k u_k.$$

If the external force is such that some of the eigenoscillations are excited very little, in other words some of the a_k are small or vanish, i.e.,

$$\sum_{k=n+1}^{N} |a_k| < \delta,$$

then it is easy to show that $\|z_n - E_n z_n\|$ is small. For

$$z_0 = g = \sum_{k=1}^{N} a_k u_k, z_1 = \sum_{k=1}^{N} a_k \lambda_k u_k, ..., z_n = \sum_{k=1}^{N} a_k \lambda_k^n u_k,$$

consider the difference

$$z_n - \overline{z_n} = Q_n(A) z_0,$$

where

$$Q_n(\lambda) = (\lambda - \lambda_1)...(\lambda - \lambda_n).$$

The fact that E_n is a projection implies that

$$\|z_n - E_n z_n\| \leq \|z_n \overline{z_n}\| = \| \sum_{k=n+1}^{N} Q_n(\lambda_k) a_k u_k \|.$$

Let M denote $\max_k |Q_n(\lambda_k)|$ and suppose that the eigenelements are normalized. Then $\|z_n - E_n z_n\|$ satisfies the following inequality

$$\|z_n - E_n z_n\| \leq \delta M.$$

Now the second case will be discussed. Assuming that some of the eigenvalues are small in modulus, i.e.,

$$|\lambda_k| < \delta, \qquad k = m+1, ..., N,$$

the difference can be formed

$$z_n - \bar{z_n} = G_n(A) z_0,$$

in which

$$G_n(\lambda) = (\lambda - \lambda_1)...(\lambda - \lambda_n)\lambda^{n-m}.$$

By the property of the projection

$$\|z_n - E_n z_n\| \leq \|z_n - \bar{z_n}\| = \| \sum_{k=m+1}^{N} \lambda_k^{n-m}(\lambda_k - \lambda_1)...(\lambda_k - \lambda_m) a_k u_k \|,$$

so

$$\|z_n - E_n z_n\| \leq \delta^{n-m}(|\lambda_1| + \delta)...(|\lambda_m| + \delta) \sum_{k=m+1}^{N} |a_k|.$$

So, $\|z_n - E_n z_n\|$ will be small when $n > m$ provided the external perturbation is such that the sum

$$\sum_{k=m+1}^{N} |a_k|$$

is not too big.

For many systems the opposite picture holds. With increasing number, the eigenvalues grow rapidly in modulus and the norm of the operator is large although bounded. The approximations converge considerably faster for such systems if instead of forming iterations with the operator A appearing in the equation (4.20) its inversion is used. The equation (4.20) can be rewritten in the following way

$$A^{-1}\frac{dx}{dt} = x + A^{-1} g f(t). \tag{4.23}$$

Then a sequence of iterations can be constructed by setting

$$z_0 = A^{-1} g, z_1 = A^{-1} z_0, ..., z_n = A^{-1} z_{n-1}.$$

Denote the operator A_n the solution of the moment problem (4.6), then the solution of the following equation is sought

$$A_n \frac{dx_n}{dt} = x_n + A^{-1} g f(t). \tag{4.24}$$

Its solution is of the form

$$z_n = \eta_0(t)z_0 + ... + \eta_{n-1}(t)z_{n-1}.$$

Substituting in the equation (4.24) and application of the formula (4.3) gives

$$-(\eta_0 + \alpha_0 \frac{d\eta_{n-1}}{dt})z_0 + (\frac{d\eta_0}{dt} - \eta_1 - \alpha_1 \frac{d\eta_{n-1}}{dt})z_1 + ...$$

$$... + (\frac{d\eta_{n-2}}{dt} - \eta_{n-1} - \alpha_{n-1} \frac{d\eta_{n-1}}{dt})z_{n-1} = z_0 f(t).$$

This equation is equivalent to the following system of scalar equations for the $\eta_k(t)$

$$-\eta_0 - \alpha_0 \frac{d\eta_{n-1}}{dt} = f(t),$$
$$\frac{d\eta_0}{dt} - \eta_1 - \alpha_1 \frac{d\eta_{n-1}}{dt} = 0,$$
$$...$$
$$\frac{d\eta_{n-2}}{dt} - \eta_{n-1} - \alpha_{n-1} \frac{d\eta_{n-1}}{dt} = 0,$$

with the initial conditions $\eta_0(0) = \eta_1(0) = ... = \eta_{n-1}(0) = 0$. It can be shown (see e.g., [41, p. 112]) that the error in the approximate solution for this case is also proportional to $\|z_n - E_n z_n\|$ which is now small for fairly small values of n since the eigenvalues of A^{-1} decreases rapidly.

Vorobyev explained a close relationship between the method of moments and the method of determining the spectrum of a matrix due to A.N.Krylov (see e.g. [21] and [22]) in his book. Krylov's starting point was also a system of differential equations with constant coefficients. The idea of the method is in the following. An arbitrary vector x is chosen in N-dimensional space and a sequence of iterations of it with the matrix A is then constructed

$$x, Ax, ..., A^k x.$$

Because of the finite dimensionality of the space, there exists n such that $A^n x$ is a linear combination of the vectors $A^k x, \quad k < n$. That is

$$A^n x = \alpha_{n-1} A^{n-1} + ... + \alpha_0 x = 0, \tag{4.25}$$

or

$$\phi(A)x = 0.$$

This implies an equation for the eigenvalues

$$\lambda^n + \alpha_{n-1}\lambda^{n-1} + ... + \alpha_0 = 0.$$

The subspace H_n generated by the vectors $x, Ax, ..., A^{n-1}x$ reduces A, and the only nth degree polynomial with leading coefficient of unity vanishing in H_n is the $P_n(\lambda)$ (4.2). The operators A and A_n coincide in H_n, and $\phi_n(\lambda) = P_n(\lambda)$. It follows that the only distinction between the method of moments and Krylov's method for the case of a finite symmetric matrix is their computational scheme.

4.5 Model reduction using the Vorobyev moment problem

In this section we will present the way how to obtain the same matching moment property as in previous Chapter (3.12) for the matrix A with the use of the Vorobyev moment problem, the presented results can be found in [38] or in the upcoming book [26].

Similarly as in previous Chapter consider a linear algebraic system $Ax = b$ with a HPD matrix A and the same initial vector x_0, giving the initial residual $r_0 = b - Ax_0$ and the Lanczos initial vector $v \equiv v_1 = r_0/\|r_0\|$.

Using the orthogonal projector $E_n = V_n V_n^*$ onto $\mathcal{K}_n(A, v_1)$ (see (3.3)), the operator A can be orthogonally restricted to the subspace $\mathcal{K}_n(A, v_1)$. Then the resulting orthogonally projected restriction A_n is given by

$$A_n = V_n V_n^* A V_n V_n^* = V_n T_n V_n^* \tag{4.26}$$

with its matrix in the orthonormal basis of $\mathcal{K}_k(A, v_1)$ represented by V_n

$$V_n^* A_n V_n = T_n.$$

The restricted operator A_n is determined by its action on the vectors generating $\mathcal{K}_k(A, v_1)$

$$\begin{aligned} A_n v_1 &= A v_1, \\ A_n(A v_1) &= A^2 v_1, \\ &\dots \\ A_n(A^{n-2} v_1) &= A^{n-1} v_1, \\ A_n(A^{n-1} v_1) &= E_n(A^n v_1) = V_n V_n^* A^n v_1. \end{aligned}$$

Equivalently,

$$\begin{aligned}A_n v_1 &= A v_1,\\ A_n^2 v_1 &= A^2 v_1,\\ &\dots\\ A_n^{n-1} v_1 &= A^{n-1} v_1,\\ A_n^n v_1 &= V_n V_n^* A^n v_1.\end{aligned}$$

In [38, p. 6] an elegant way how to get the same matching moment property as (3.12) is shown. By construction

$$v_1^* A^k v_1 = v_1^* A_n^k v_1, \quad k = 1, 2, ..., n. \tag{4.27}$$

Since $\mathcal{K}_n(A, v_1) = \text{span}\{v_1, Av_1, ..., A^{n-1}v_1\}$ and $(A_n)^n v_1 \in \mathcal{K}_k(A, v_1)$, the orthogonal projection

$$0 = E_n(A^n v_1) - (A^n) v_1 = E_n(A^n v_1 - (A_n)^n v_1)$$

implies that the difference $A^n v_1 - (A_n)^n v_1$ must be orthogonal to all basis vectors $v_1, Av_1 = A_n v_1, ..., A^{n-1} v_1 = A_n^{n-1} v_1$, which gives, using the properties $A^* = A$, $A_n^* = A_n$,

$$0 = (A^j v_1) * (A^n v_1 - (A_n)^n v_1) = v_1^* A^{n+j} v_1 - v_1^* (A_n)^{n+j} v_1, \quad j = 0, 1, ..., n-1.$$

Combining with (4.27) and (4.26) this gives

$$v_1^* A^k v_1 = v_1^* A_n^k v_1 = e_1^T T_n^k e_1, \qquad k = 0, 1, ..., 2n-1,$$

which is the same matching moment property as (3.12).

In [38, pp. 7-11] the proofs of the similar identities for A non-Hermitian are given. Let's start with the algorithm of non-Hermitian Lanczos process. Given a non-singular matrix $A \in \mathbb{C}^{N \times N}$ and two starting vectors $v \equiv v_1, w \equiv w_1$ of length N, $\|v_1\| = 1, w_1^* v_1 = 1$, the non-Hermitian Lanczos algorithm can be written in the form

$$AV_n = V_n T_n + \delta_{n+1} v_{n+1} e_n^T,$$

$$A^* W_n = W_n T_n^* + \beta_{n+1}^* w_{n+1} e_n^T,$$

where $W_n^* V_n = I, T_n = W_n^* A V_n, \|v_{n+1}\| = 1, w_{n+1}^* v_{n+1} = 1$,

$$T_n = \begin{bmatrix} \gamma_1 & \beta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \beta_n \\ & & \delta_n & \gamma_n \end{bmatrix}, \quad \delta_l > 0, \beta_l \neq 0, l = 2, ...n. \tag{4.28}$$

Here it is assumed that the algorithm does not break down in any step. The columns of V_n form a basis of $\mathcal{K}_n(A, v_1)$, the columns of W_n form a basis of $\mathcal{K}_n(A^*, w_1)$. Because of the biorthogonality $W_n^* V_n = I$, the oblique projector onto $\mathcal{K}_n(A, v_1)$ orthogonal to $\mathcal{K}_n(A^*, w_1)$ can be written as

$$Q_n = V_n W_n^*.$$

In [38, pp. 7-9], a proof, that the non-Hermitian Lanczos algorithm represents the model reduction which matches the first $2n$ moments is given

$$w_1^* A^k v_1 = e_1^T T_n^k, e_1, \quad k = 0, 1, ..., 2n-1.$$

The proof is based on the relationship with the corresponding Vorobyev moment problem. It is similar to the proof of the matching moment property for the Lanczos but it is technically a little bit more difficult.

Now let's show a matching moment property for the Arnoldi method. Given a non-singular matrix $A \in \mathbb{C}^{N \times N}$ and an initial vector $v \equiv v_1$ of length N, $\|v_1\| = 1$, the Arnoldi algorithm can be seen as

$$AV_n = V_n H_n + h_{n+1,n} v_{n+1} e_n^T,$$

where

$$V_n^* V_n = I_n, \quad H_n = V_n^* A V_n,$$

where H_n is upper Hessenberg matrix with positive entries on the first subdiagonal. In [38, pp. 10-11] a proof of the following matching moment property for the Arnoldi method is given

$$v_1^* A^k v_1 = e_1^T H_n^k e_1, \quad k = 0, 1, ..., n.$$

So this model reduction matches $n+1$ moments. In the proof the Vorobyev moment problem is again used in order to show the matching moment property.

Chapter 5

Relationship to the Sturm-Liouville problem

Sturm-Liouville problems are the problems which involves solving the second order ordinary differential equations. Solution of equations of these types has played a fundamental role in the evolution of mathematics and physics, starting with the eigenoscillations of a string and culminating in the atomic vibrations of Shrödinger's wave equation. In [25, pp. 348-432] one can see a brief introduction of the history of Sturm-Liouville problems and solutions of many special types of second order ordinary differential equations. The aim of this Chapter is to show the connection between the moment problem and one type of the Sturm-Liouville equations. Consider the Sturm-liouville problem in the following form

$$-y^{''} + q(x)y = \lambda y, \quad x \in [a, b], \tag{5.1}$$

where a, b may be finite or infinite and appropriate boundary conditions for y are given. The Sturm-Liouville problem is called *regular* if a and b are finite, $p(x)$ is continuous except for finitely many jumps and regular boundary conditions are set, i.e.,

$$y(a) = c_1 y^{'}(a), \quad y(b) = c_2 y^{'}(b), \quad c_1, c_2 \in \mathbb{R}.$$

The equation (5.1) can be written in the following form

$$Ly = \lambda y,$$

where L is a linear differential operator defined as follows

$$L \equiv -\frac{d^2}{dx^2} + q(x).$$

In the case of the regular Sturm-Liouville problem the solutions of (5.1) form a complete orthogonal set in $L_2(a, b)$. For a proof and for many more interesting properties of the solutions of Sturm-Liouville problems see e.g., [39].

Many books and articles have been written about Sturm-Liouville problems. More information about the classification and solving the Sturm-Liouville problems can be found e.g. in [32], [11] or [25].

In [12] the connections between the *singular* Sturm-Liouville problem, Jacobi matrices and Hamburger moment problem are described in an elegant way. In this Chapter we will present some of the results. Consider the *singular* Sturm-Liouville boundary value problem

$$-y^{''} + q(x)y = \lambda y, \quad x \in [0, \infty), \tag{5.2}$$

with boundary condition

$$y(0)\cos\alpha + y^{'}(0)\sin\alpha = 0,$$

where $q(x)$ is a continuous function on $[0, \infty)$. Now let $\phi(x, \lambda)$, $\theta(x, \lambda)$ be the solutions of the differential equation of the boundary value problem (5.2) such that

$$\phi(0, \lambda) = \sin\alpha, \quad \phi^{'}(0, \lambda) = -\cos\alpha,$$

$$\theta(0, \lambda) = \cos\alpha, \quad \theta(0, \lambda) = \sin\alpha.$$

In [39] it is shown, that there exists a complex valued function $m_\infty(\lambda)$, the so-called Weyl-Titchmarsh function, such that for every $\lambda \in \mathbb{C}\backslash\mathbb{R}$ the differential equation (5.2) has a solution

$$\chi_\infty(x, \lambda) = \theta(x, \lambda) + m_\infty(\lambda)\phi(x, \lambda),$$

belonging to $L_2(0, \infty)$. In the so-called limit-point case $m_\infty(\lambda)$ is unique, while in the limit-circle case uncountably many such functions exists. For each $\lambda \in \mathbb{C}\backslash\mathbb{R}$, the values of these functions belong to a geometrical circle $C_\infty(\lambda)$, in the limit-point case, this circle collapses into a point. Here the index ∞ in m_∞ is connected with the properties of the function $q(x)$ in (5.2) which is supposed only to be continuous on $[0, \infty)$ and so it could not be integrable through the whole $[0, \infty)$.

In the same way the Sturm-Liouville difference equation can be considered

$$\nabla(\delta_{n+1}\Delta z_{n-1}) + (\gamma_{n+1} + \delta_{n+1} + \delta_n)z_{n-1} = \lambda z_{n-1}, \quad n = 1, 2, ...$$

where ∇ and Δ denotes backward and forward differences. This difference equation can be written in the following form

$$\delta_{n+1}z_n + \gamma_n z_{n-1} + \delta_n z_{n-2} = \lambda z_{n-1}, \quad n = 1, 2, ..., \tag{5.3}$$

where $\delta_1 = 0, \delta_n > 0$ and $\gamma_n \in \mathbb{R}$ for each $n \in \mathbb{N}$ along with the initial data $z_{-1} = 0, z_0 = 1$ and $z_{-1} = -1, z_0 = 0$. This is again the three-term Stieltjes recurrence for orthogonal polynomials, see (3.9).

In an analogic way as in the continuous case, consider the following solutions of the difference equation (5.3) for both initial data.

$$\Phi(\lambda) = (\varphi_0(\lambda), \varphi_1(\lambda), \varphi_2(\lambda), ...), \qquad \Psi(\lambda) = (\psi_0(\lambda), \psi_1(\lambda), \psi_2(\lambda), ...),$$

and the Weyl-Titchmarsh function is searched in such a way that

$$\chi_\infty(\lambda, n) = \Phi(\lambda) + m_\infty(\lambda)\Psi(\lambda) \in l^2(\mathbb{N}_0)$$

is the solution of the difference equation (5.3). If there exists an unique function $m_\infty(\lambda)$, we are in the limit-point case, while in the limit-circle case uncountably many such functions exist.

Now consider the following semi-infinite jacobi matrix T which defines an operator on $l_2(\mathbb{N}_0)$ associated with the difference equation (5.3)

$$T = \begin{bmatrix} \gamma_1 & \delta_2 & 0 & 0 & \dots \\ \delta_2 & \gamma_2 & \delta_3 & 0 & \dots \\ 0 & \delta_3 & \gamma_3 & \delta_4 & \ddots \\ 0 & 0 & \delta_4 & \gamma_4 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{5.4}$$

It is known, that there is a close relationship between problem (5.3) being in the limit-point or in the limit-circle case and the existence of self-adjoint extensions of T (5.4). Both problems are equivalent to deciding the determinacy of the Hamburger moment problem associated with T (5.4).

Given the real numbers $\xi_k = (\delta_0, T^k\delta_0)_{l^2}, k \geq 0$, where δ_0 stands for the sequence $(1, 0, 0, ...)$, we are interested in the search of positive measures μ such that

$$\int_{-\infty}^{\infty} x^k d\mu(x) = \xi_k, \qquad k = 0, 1, ... \tag{5.5}$$

The above problem is called the Hamburger moment problem associated with T (5.4). Denote S a self-adjoint extension of T (see basic notation in the beginning of this thesis), i.e.,

$$T \subset S \subset T^*$$

In [34, p. 5, pp. 21-22] a proof of the following theorem which characterizes connection between the existence of self-adjoint extensions of T (5.4) and the determinacy of the associated Hamburger moment problem (5.5) is given.

Theorem 5.0.1 *A necessary and sufficient condition that the measure $d\mu$ in (5.5) be unique is that the operator T of (5.4) is essentially self-adjoint, i.e., has a unique self-adjoint extension.*

It means, that T is not an essentially self-adjoint operator if and only if the associated moment problem is indeterminate. Then its self-adjoint extensions S_t can be parametrized by $t \in \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ and as shown in [34, p. 19] or [12, p. 4] their domains are

$$\mathcal{D}(S_t) = \begin{cases} \mathcal{D}(T) + \text{span}\{t\Phi(0) + \Psi(0)\} & \text{if} \quad t \in \mathbb{R}, \\ \mathcal{D}(T) + \text{span}\{\Phi(0)\} & \text{if} \quad t = \infty. \end{cases}$$

In [34, pp. 38-39] it is shown that each self-adjoint extension of T has a pure point spectrum $\{\lambda_i = \lambda_i(S_t)\}_{i=0}^{\infty}$. The corresponding eigenfunctions $\{\Phi_i = \Phi(\lambda_i)\}_{i=0}^{\infty}$. are given by

$$\Phi_i = (\varphi_0(\lambda_i), \varphi_1(\lambda_i), \varphi_2(\lambda_i), ...) \quad i \in \mathbb{N}_0,$$

and they form an orthogonal basis in $l^2(\mathbb{N}_0)$.

Now let $\{\xi_k\}_{k=0}^{\infty}$ be an indeterminate Hamburger moment sequence associated with T (5.4). Let's denote by V the set of positive measures satisfying it, i.e.,

$$V = \left\{ \mu \geq 0 : \xi_k = \int_{-\infty}^{\infty} x^k d\mu(x), k \geq 0 \right\}.$$

The moment problem is related with the self adjoint extensions of the semi-infinite Jacobi matrix T. Let $\{\lambda_i^t = \lambda_i(S_t)\}_{i=0}^{\infty}$ be the eigenvalues associated with the self-adjoint extensions S_t of T. The set V can be parametrized (see e.g., [34, p. 12]) in order to obtain the set of Neumann measures $\{\mu_t\}$ which satisfy

$$(\delta_0, (\lambda I - S_t)^{-1}, \delta_0) = \int_{-\infty}^{\infty} \frac{d\mu_t(x)}{\lambda - x} = \frac{A_{Nev}(\lambda) + tC_{Nev}(\lambda)}{B_{Nev}(\lambda) + tD_{Nev}(\lambda)}, \qquad \lambda \in \mathbb{C} \backslash \mathbb{R},$$

where $A_{Nev}, B_{Nev}, C_{Nev}, D_{Nev}$ are the functions of the so-called Nevanlinna matrix associated with the moment problem. In [34, p. 36] the explicit formulas of the Nevanlinna matrix are given. It is known (see e.g., [12, p. 8]), that for each $t \in \overline{\mathbb{R}}, \mu_t$ is the discrete measure $\mu_t = \sum_{\lambda \in Z_t} a_\lambda \delta_\lambda$, where

$$Z_t = \begin{cases} \{\lambda \in \mathbb{C} | B_{Nev}(\lambda) + tD_{Nev}(\lambda) = 0\} & \text{if} \quad t \in \mathbb{R}, \\ \{\lambda \in \mathbb{C} | D_{Nev}(\lambda) = 0\} & \text{if} \quad t = \infty, \end{cases}$$

and

$$a_\lambda = \frac{A_{Nev}(\lambda) + tC_{Nev}(\lambda)}{B'_{Nev}(\lambda) + tD'_{Nev}(\lambda)}, \qquad \lambda \in Z_t.$$

The measure μ_t is then the spectral measure of S_t. In [34, pp. 38-39] it is shown, that the zeros of the function $B_{Nev}(\lambda) + tD_{Nev}(\lambda)$ (or of the $D_{Nev}(\lambda)$) which are real and simple, are precisely the sequence $\{\lambda_i^t\}_{i=0}^{\infty}$ and moreover

$$\mu_t(\{\lambda_i^t\}) = 1/\|\Pi_i\|^2.$$

It is worth to say that Akhiezer [1] calls these Neumann measures N-extremal and in Shoat-Tamarkin's book [33] they are called extremal.

In [34, p. 41] and [12, p. 8] there can be found the following characterization for $m_\infty(\lambda)$. If $C_\infty(\lambda)$ is the limit circle associated to $\lambda \in \mathbb{C}\backslash\mathbb{R}$, then $m_\infty(\lambda) \in C_\infty(\lambda)$ if and only if it admits the following parametrization

$$m_\infty^t(\lambda) = (\delta_0, (\lambda I - S_t)^{-1}, \delta_0) = \int_{-\infty}^{\infty} \frac{d\mu_t(x)}{\lambda - x} = \frac{A_{Nev}(\lambda) + tC_{Nev}(\lambda)}{B_{Nev}(\lambda) + tD_{Nev}(\lambda)}, \quad t \in \overline{\mathbb{R}},$$

where $A_{Nev}, B_{Nev}, C_{Nev}, D_{Nev}$ are the components of the Nevalinna matrix and μ_t is the Neumann spectral measure associated with S_t. .

So, the Hamburger moment problem (5.5) associated with T (5.4) is indeterminate if and only if the difference equation (5.3) belongs to the limit-circle case. This provides a criteria to decide when a difference Sturm-Liouville problem (5.3) belongs to the limit-circle case.

Chapter 6

Model reduction of large-scale dynamical systems

In this Chapter we will show the connections of the model reduction of the linear dynamical systems, modern iterative methods and the moment problem. Linear dynamical system can be represented in the following form.

$$\begin{aligned} z^{'}(t) &= Az(t) + bu(t), \\ y(t) &= b^{*}z(t), \end{aligned} \tag{6.1}$$

where $A \in \mathbb{C}^{N \times N}, b \in \mathbb{C}^{N}$, are given, $z(t) \in \mathbb{C}^{N}$ represents the state of the system at time t and $u(t)$ and $y(t)$ represents scalar input and output of the system (6.1). Close relation between the theory of linear dynamical systems and modern iterative methods is known. We will define the controllability for the dynamical system (6.1). The dynamical system (6.1) is said to be controllable if for every initial condition $z(0)$ and every vector $z_1 \in \mathbb{C}^{N}$, there exists a finite time t_1 and control $u(t) \in \mathbb{C}, t \in [0, t_1]$, such that $z(t_1, z(0), u) = z_1$.
For simplicity, consider a linear steady discrete system with the state equation

$$z_{k+1} = Az_k + bu_k \tag{6.2}$$

The solution of this equation is

$$z_k = A^k z_0 + \sum_{i=0}^{k-1} A^{k-i-1} b u_i,$$

with

$$z_0 = 0, z_1 = bu_0.$$

In [17, p.143] the proof the following theorem is given

Theorem 6.0.2 *Linear steady discrete system (6.2) is controllable if and only if, when the matrix R*

$$R = [b, Ab, ..., A^{N-1}b]$$

has the rank N.

This is equivalent with the case, when the Krylov space $\mathcal{K}_N(A, b)$ (see (3.2)) has full rank N. Which is according to [2] equivalent to the fact that the GMRES method for system

$$Ax = b,$$

with the initial vector b doesn't stop until the last step N. So, this is example of the link between the controllability of the simple dynamical system with the properties of the method for solving systems of linear algebraic equations.

Now let's go back to the system (6.1). In [26, pp. 101-108] an elegant description of the connection between the model reduction of this system and the moment problem is given. Applying the Laplace transform,

$$\hat{f}(\lambda) := \int_0^\infty f(t)e^{-\lambda t}dt$$

the system (6.1) can be represented by the transfer function description

$$T(\lambda) = b^*(\lambda I - A)^{-1}b, \quad \lambda \in \mathbb{C}, \tag{6.3}$$

where

$$T(\lambda) = \frac{\hat{y}(\lambda)}{\hat{u}(\lambda)}.$$

The model reduction problem is to find the reduced order A_n, b_n such that

$$T_n(\lambda) = b_n^*(\lambda I - A_n)^{-1}b_n, \quad \lambda \in \mathbb{C},$$

approximates in some sense well $T(\lambda)$ within a given frequency of range $\lambda \in \mathbb{C}_A \in \mathbb{C}$. The double (A_n, b_n) is called a realization of T. In [31, p. 9] it is shown, that in certain cases it is possible to have a double (A_n, b_n) with $A_n \in \mathbb{C}^{n \times n}$ and $n < N$ which correspond to the same transfer function T as does (A, b). Such a realization with additional condition that A_n has minimal dimension is called a *minimal realization.*

The problem of finding efficient numerical approximation to (6.3) arises in many applications unrelated to linear dynamical systems (6.1). A more general case can be written as

$$c^*F(A)b,$$

where $F(A)$ is a given function of the matrix A. The particular case $c = b$ and $F(A) = (\lambda I - A)^{-1}$, i.e., $F(A)$ is equal to the matrix resolvent, where λ is outside of the spectrum A, is of great importance. Model reduction in linear dynamical systems based on projections onto Krylov subspaces is linked with local approximation of the transfer function $T(\lambda)$. First consider the expansion about infinity

$$\begin{aligned} -T(\lambda) &= \lambda^{-1} b^*(I - \lambda^{-1}A)^{-1} b = \\ &= \lambda^{-1}(b^*b) + \lambda^{-2}(b^*Ab) + ... + \lambda^{-2n}(b^*A^{2n-1}b) + ... \end{aligned}$$

A reduced model of order n which matches the first $2n$ terms in the above expansion is known as the *minimal partial realization*.

In order to see the link, consider the distribution function $\omega(\lambda)$ with N points of increase associated with the HPD matrix A and the initial vector b. Then

$$b^*(\lambda I - A)^{-1} b = \sum_{j=1}^{N} \frac{\omega_j}{\lambda - \lambda_j} = \mathcal{F}_N(\lambda),$$

where the continued fraction $\mathcal{F}_N(\lambda)$ can be for any $n < N$ expanded to

$$\mathcal{F}_N(\lambda) = \sum_{l=1}^{2n} \frac{\xi_{l-1}}{\lambda^l} + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}}\right) = \mathcal{F}_n(\lambda) + \mathcal{O}\left(\frac{1}{\lambda^{2n+1}}\right),$$

$$\xi_{l-1} = \int_a^b \lambda^{l-1} d\omega(\lambda) = \sum_{j=1}^{n} \omega_j^{(n)} \{\lambda_j^{(n)}\}^{l-1}, \quad l = 1, 2, ..., 2n,$$

or in the matrix form

$$\xi_{l-1} = b^* A^{l-1} b = e_1^T T_n^{l-1} e_1.$$

$\mathcal{F}_n(\lambda)$ approximates $\mathcal{F}_N(\lambda)$ with the error proportional to $1/\lambda^{2n+1}$. The minimal partial realization in model reduction of linear dynamical systems matches the first $2n$ moments

$$\mu_{-l} = b^* A^{l-1} b, \quad l = 1, ..., 2n,$$

called in the dynamical systems literature Markov parameters.

Sometimes it is more convenient to do the model reduction with the expansion of the $T(\lambda)$ in the neighborhood of some $\lambda_0 \in \mathbb{C}$, see e.g., [10]. Here, the case where $\lambda_0 = 0$ is used. The model reduction is achieved by matching the moments

$$\mu_l = b^*(A^{-1})^l b, \quad l = 1, 2, ...$$

of the expansion

$$-T(\lambda) = b^* A^{-1}(I - \lambda A^{-1})^{-1} b =$$
$$= b^* A^{-1} b + \lambda(b^* A^{-2} b) + ... + \lambda^{2n-1}(b^* A^{-2n} b) + ...$$

It should be noted that minimal partial realization in the theory of large scale dynamical systems was introduced by Kalman [20] in 1979. But it can be seen above that the ideas of this methods are much older and originates in the works of Chebyshev [7] and Stieltjes [35], see Chapter 3.

An interesting case is when we take $\lambda = 0$ and the new vector $c \in \mathbb{C}^N$ in (6.3). We get the following quantity

$$c^* A^{-1} b.$$

The approximation of this quantity, called in signal processing the scattering amplitude, is very important in many applications, see e.g., [14] and [15]. However as pointed out in [36] problem of numerical approximation of the single scalar value $c^* A^{-1} b$ is different from the numerical approximation of the whole transfer function $T(\lambda)$ and therefore a different approach must be taken. The approach to approximate this quadratic form $c^* A^{-1} b$ was taken in the paper [37] and a Vorobyev moment problem is used in order to get good approximation.

Chapter 7

Numerical illustrations

The problem of moments can be seen as the theoretical background for many numerical methods and therefore an insight through the moments can help to bring some new knowledge about these methods. In the Chapter 3 we have shown, how the Gauss-Christoffel quadrature can be seen as a matching moment model reduction. In [30] the results about sensitivity of Gauss-Christoffel quadrature with respect to small perturbations of the distribution function are given. Obtaining of these results would not be possible without the deep knowledge of the connection with the moment problem.

Consider a sufficiently smooth function $f(\lambda)$ uncorrelated with the perturbation of the distribution function. It seems natural that the difference between the quadrature approximates is of the same order as the difference between the original and perturbed integral. As pointed out in [30, p. 1] this is one of the reasons why nobody formulated this question in literature before 2007 when the article [30] came out.

The nodes of the n-point Gauss-Christoffel quadrature are the roots nth orthogonal polynomial associated with the distribution function. The perturbed distribution function generates different sequence of orthogonal polynomials and this fact may cause that the difference between the quadratures could be of the higher order than the difference between the integrals.

In [30] a characterization of the case, when the quadrature is insensitive to the small perturbation of the distribution function is given. It is when the size of the support of the perturbed distribution function remains the same. In other way, when the perturbation changes the size of the support of the distribution function, the quadrature may be very sensitive.

The question whether the condition numbers of the matrices of modified and mixed moments (these matrices will be precisely defined in the following text), which are nothing else than matrices of moments Δ_k from the Chapter 2 (see (2.6) and (2.7)), could help us to decide when the perturbation causes the big difference in quadrature estimates and when these estimates remains insensitive is also posed in [30]. The questions about the conditioning of these matrices were studied by Gautschi [13] and by Beckermann and Bourreau in [4]. They showed that condition numbers of matrices of modified and mixed moments grow exponentially if the support of the original and perturbed distribution function changes, see e.g. [30, p. 23] and [4, p. 93]. So, though it would be natural that in the case of sensitivity, the matrices of the moments would be ill-conditioned and in the other case well-conditioned, [30, p. 23]. Due to the nature of matrices of moments, which are called Hankel matrices, this presumption doesn't hold. These matrices seems to be ill-conditioned in both cases.

The aim of this Chapter is to show several numerical illustrations with slightly different parameters than in [30] in order to show ill-conditioning of the Hankel matrices. The same software package in MATLAB for computing the Gauss-Christoffel quadrature, which was used in [30] is used here.

Let $I_\Omega(f)$ be the Riemann-Stieltjes integral of the function f with the non-decreasing distribution function $\Omega(\lambda)$

$$I_\Omega(f) = \int_a^b f(\lambda) d\Omega(\lambda)$$

we wish to approximate it by the n-point Gauss-Christoffel quadrature

$$I_\Omega^n(f) = \sum_{l=1}^{n} \Omega_l^{(n)} f(\lambda_l^{(n)}).$$

Let $f = x^{-1}$ and consider the points of increase $0 < a < \lambda_1 < \lambda_2 < ... < \lambda_n < b$,

$$\lambda_l = \lambda_1 + \frac{l-1}{n-1}(\lambda_n - \lambda_1)\gamma^{n-l}, \quad \gamma \in (0,1), \quad l = 2, 3, ..., n-1,$$

and randomly generated jumps $\Omega_1, \Omega_2, ..., \Omega_n$ which are normalized such that

$$\sum_{l=1}^{n} \Omega_l = 1,$$

Consider the distribution function $\Phi(\lambda, \sigma) = \sum_{l=1}^{n} \Omega_l \varphi(\lambda; \sigma, \lambda_l)$, where $\varphi(\lambda; \sigma, t)$ is continuous piecewise linear function on $[t-\sigma, t+\sigma]$, which is linear on $[t-\sigma, t]$ and on $[t, t+\sigma]$, and is constant on $(-\infty, t-\sigma]$ and on $[t+\sigma, \infty)$.
In our examples we will use the distribution function

$$\Omega(\lambda, \sigma) = c\Phi(\lambda, \sigma), \quad \int_a^b d\Omega(\lambda, \sigma) = 1,$$

where c is the normalization constant. We will use $\sigma = 10^{-8}$ for the original distribution function $\Omega_0 \equiv \Omega(\lambda, 10^{-8})$ and $\sigma = 10^{-6}$ for the perturbed distribution function $\Omega_1 \equiv \Omega(\lambda, 10^{-6})$. We set $n = 30$, $a = \lambda_1 - 10^5 = 10^{-1} - 10^{-5}$, $b = \lambda_n + 10^{-5} = 500 + 10^{-5}$.

As the example of the case where the size of the support of the distribution function remains the same we use the distribution function $\Omega_2 \equiv \Omega(\lambda, 10^{-8})$ where the intervals of increase $[\lambda_l - \sigma, \lambda_l + \sigma]$ are shifted randomly to the left or right while the difference between Ω_2 and Ω_0 remains of the same order as the difference between Ω_0 and Ω_1. The components of the matrices of modified and mixed moments GM_n and MM_n are determined by the inner product of the orthogonal polynomials with respect to the corresponding distribution functions

$$GM_n(i,j) = (\psi_i, \psi_j),$$
$$MM_n(i,j) = (\varphi_i, \psi_j),$$

where the inner product is determined by the original distribution function Ω_0, φ_k are the orthogonal polynomials with respect to the original distribution function Ω_0 and ψ_k are the orthogonal polynomials with respect to the perturbed distribution functions Ω_1 or Ω_2.
We will present numerical illustrations for $\gamma = 0.2, 0.4, 0.6, 0.8$.

$\gamma = 0.2$

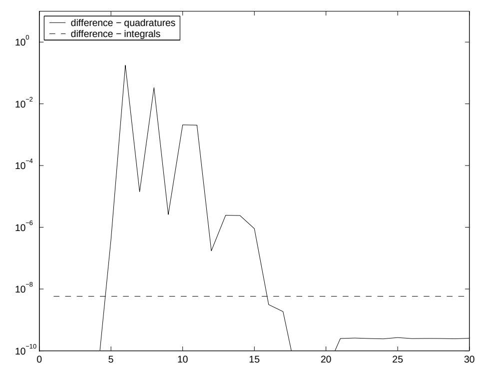

Figure 7.1: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_1

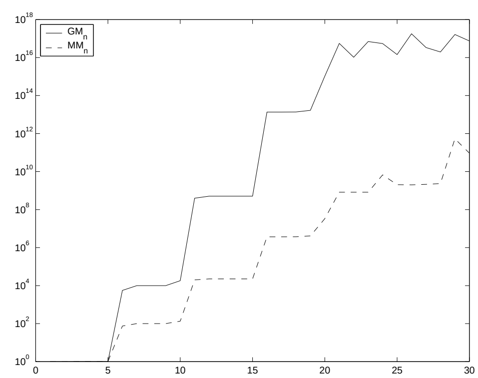

Figure 7.2: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_1

$\gamma = 0.2$

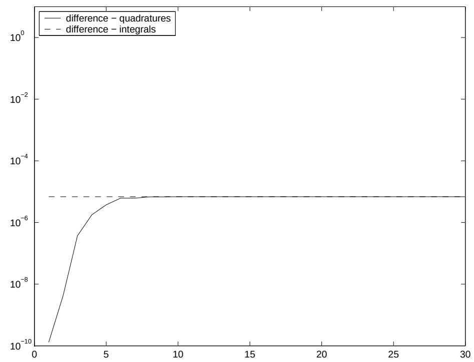

Figure 7.3: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_2

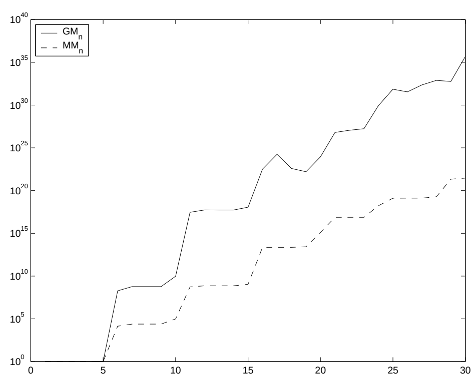

Figure 7.4: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_2

$\gamma = 0.4$

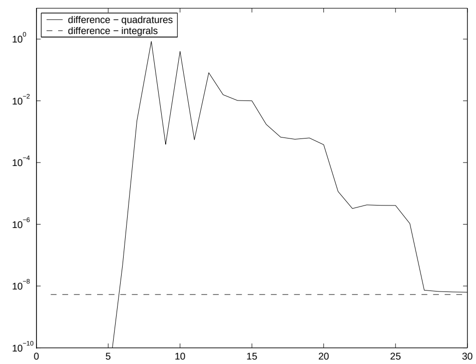

Figure 7.5: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_1

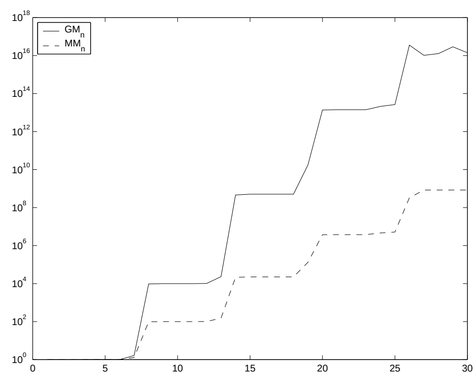

Figure 7.6: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_1

$\gamma = 0.4$

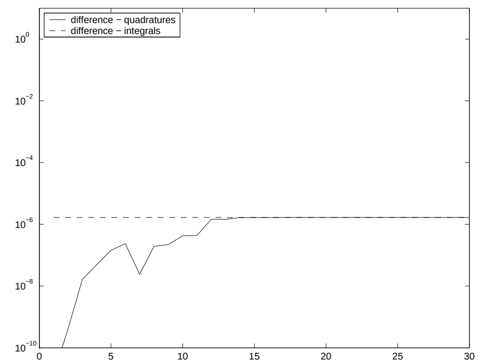

Figure 7.7: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_2

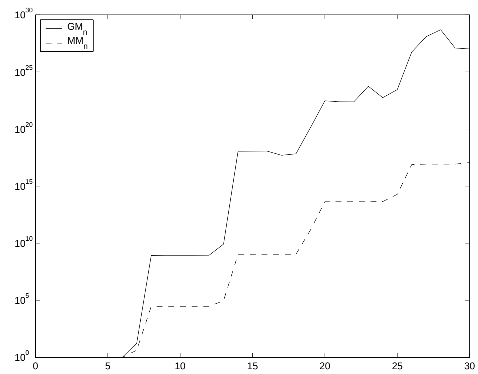

Figure 7.8: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_2

$\gamma = 0.6$

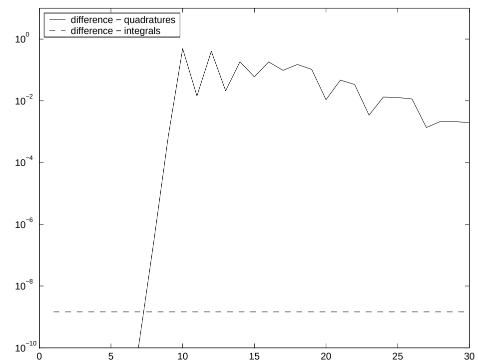

Figure 7.9: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_1

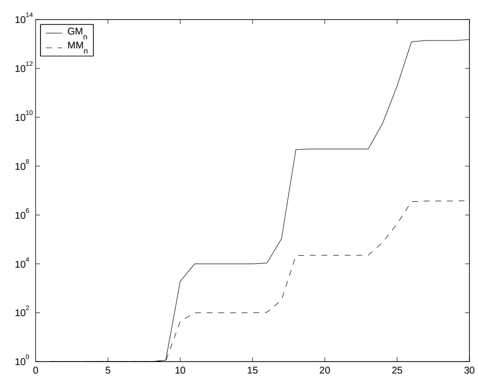

Figure 7.10: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_1

$\gamma = 0.6$

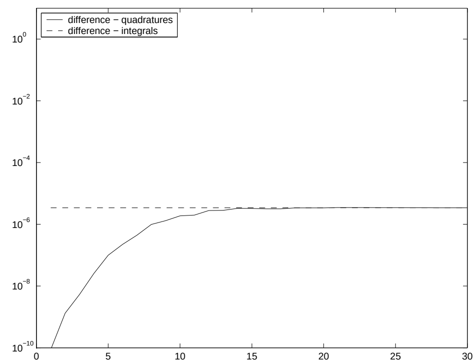

Figure 7.11: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_2

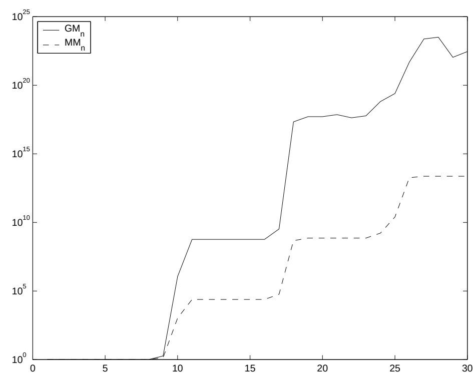

Figure 7.12: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_2

$\gamma = 0.8$

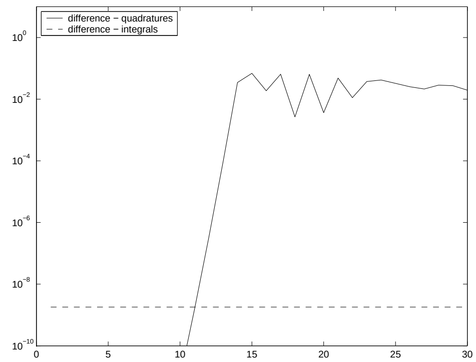

Figure 7.13: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_1

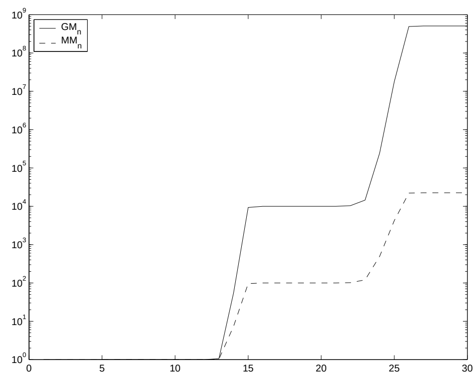

Figure 7.14: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_1

$\gamma = 0.8$

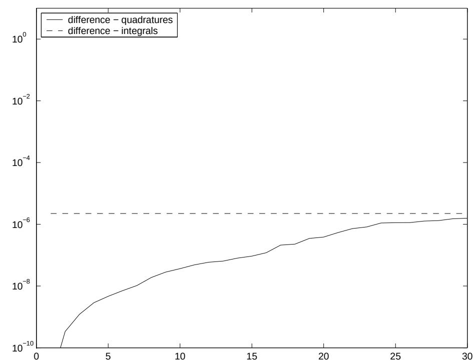

Figure 7.15: Absolute value of the difference for the quadratures and integrals corresponding to Ω_0 and Ω_2

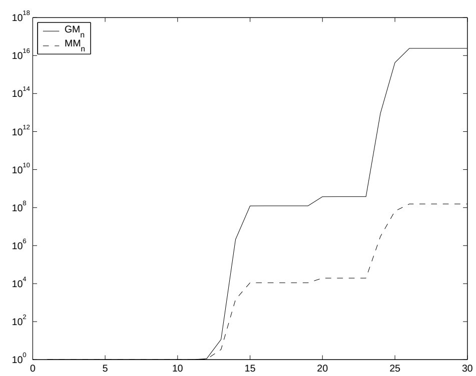

Figure 7.16: Condition numbers of GM_n and MM_n corresponding to the distribution functions Ω_0 and Ω_2

As we can see in the above figures, the Hankel matrices GM_n and MM_n are ill-conditioned in both discussed cases of perturbation of distribution function. So, their condition numbers doesn't give us any information whether the Gauss-Christoffel quadrature will be sensitive to the perturbation.

Chapter 8

Conclusion

In this work we have shown some connections between different mathematical approaches through the problem of moments. In the Chapter 2 a brief historical review of the study of the moment problem was given. In the Chapters 3 and 4 two different approaches to the moment problem were studied in order to show the connection with the modern iterative methods. In the Chapters 5 and 6 it was shown how the moment problem is connected with the solving of equations of Sturm-Liouville type and with the solution of large scale dynamical systems. It was shown, that the moment problem can be seen as the theoretical background in many mathematical methods, especially in the Gauss-Christoffel quadrature, Lanczos method and CG method. The deep knowledge of the mechanisms behind the numerical methods could lead to the new results as shown in the Chapter 7 devoted to the numerical illustrations. So, the study of the connections through the moment problem could be very useful and could reveal new ways in the research. The contribution of this work is the summarizing of these connections in an understandable way.

Bibliography

[1] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.

[2] M. Arioli, V. Pták, and Z. Strakoš. Krylov sequences of maximal length and convergence of GMRES. *BIT*, 38(4):636–643, 1998.

[3] W. V. Assche. The impact of stieltjes' work on continued fractions and orthogonal polynomials. *J. Comput. Appl. Math.*, 65:419–447, 1995.

[4] B. Beckermann and Bourreau E. How to choose modified moments? *J. Comput. Appl. Math*, 98(1):81–98, 1998.

[5] C. Brezinski. *History of continued fractions and Padé approximants*, volume 12 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1991.

[6] P. Butzer and F. Jongmans. P. L. Chebyshev (1821–1894). A guide to his life and work. *J. Approx. Theory*, 96(1):111–138, 1999.

[7] P. Chebyshev. Sur les fractions continues. *Oeuvres I.*, 11, 1855.

[8] P. Chebyshev. Sur les valeurs limites des integrales. *Journal des Mathématiques pures et appliquées*, 19, 1874.

[9] P. J. Davis and P. Rabinowitz. *Methods of numerical integration*. Computer Science and Applied Mathematics. Academic Press Inc., Orlando, FL, second edition, 1984.

[10] P. V. Dooren, K. Gallivan, and Grimme E. Asymptotic waveform evaluation via a lanczos method. *Appl. Math. Lett.*, 7:75–80, 1994.

[11] F. P. Gantmacher and M. G. Krein. *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*. American Mathematical Society, 2002.

[12] A. G. García and M. A. Hernández-Medina. Discrete Sturm-Liouville problems, Jacobi matrices and Lagrange interpolation series. *J. Math. Anal. Appl.*, 280(2):221–231, 2003.

[13] W. Gautschi. Orthogonal polynomials: computation and approximation. *Numerical Mathematics and Scientific Computation. Oxford University Press*, 2004.

[14] G. H. Golub and G. Meurant. Matrices, moments and quadrature. In *Numerical analysis 1993 (Dundee, 1993)*, volume 303 of *Pitman Res. Notes Math. Ser.*, pages 105–156. Longman Sci. Tech., Harlow, 1994.

[15] G. H. Golub and Z. Strakoš. Estimates in quadratic formulas. *Numer. Algorithms*, 8(2-4):241–268, 1994.

[16] H. Hamburger. Über eine Erweiterung des Stieltjesschen Momentenproblems. *Math. Ann.*, 81(2-4):235–319, 1920.

[17] V. Havlena and J. Štecha. *Teorie dynamických systémů.* ČVUT, 2005.

[18] E. Heine. *Handbuch der Kugelfunctionen. Theorie und Anwendungen. Band I, II.* Zweite umgearbeitete und vermehrte Auflage. Thesaurus Mathematicae, No. 1. Physica-Verlag, Würzburg, 1861, 1861.

[19] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.

[20] R. E. Kalman. On partial realizations, transfer functions, and canonical forms. *Acta Polytech. Scand. Math. Comput. Sci. Ser.*, 31:9–32, 1979. Topics in systems theory.

[21] A. N. Krylov. On the numerical solution of equations, which determine the oscillation frequencies of physical systems in engineering problems. *Izv. Akad. Nauk SSSR*, pages 491–539, 1931.

[22] A. N. Krylov. *On some differential equations of mathematical physic.* Akad. Nauk SSSR, 1933.

[23] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Research Nat. Bur. Standards*, 45:255–282, 1950.

[24] C. Lanczos. Solution of systems of linear equations by minimized-iterations. *J. Research Nat. Bur. Standards*, 49:33–53, 1952.

[25] C. Lanczos. *Linear differential operators*. D. VAN NOSTRAND COMPANY, LTD., 1961.

[26] J. Liesen and Z. Strakoš. *Principles and Analysis of Krylov Subspace Methods*. In preparation, 2010.

[27] A. Markov. Démonstration de certaines inégalités de M. Tchébychef. *Math. Ann.*, 24(2):172–180, 1884.

[28] A. Markov. Sur une question de maximum et de minimum. *Acta Math.*, 9(1):57–70, 1887. Proposée par M. tchebycheff.

[29] G. Meurant and Z. Strakoš. The Lanczos and conjugate gradient algorithms in finite precision arithmetic. *Acta Numer.*, 15:471–542, 2006.

[30] D. P. O'Leary, Z. Strakoš, and P. Tichý. On sesitivity of gauss-christoffel quadrature. *Springer-Verlag*, 2007.

[31] B. N. Parlett. Reduction to tridiagonal form and minimal realizations. *SIAM J. Matrix Anal. Appl.*, 13(2):567–593, 1992.

[32] J. D. Pryce. *Numerical Solution of Sturm-Liouville Problems*. Oxford science publications, 1993.

[33] J. A. Shohat and J. D. Tamarkin. *The Problem of Moments*. American Mathematical Society Mathematical surveys, vol. II. American Mathematical Society, New York, 1943.

[34] B. Simon. The classical moment problem as a self-adjoint finite difference operator. *Adv. Math.*, 137(1):82–203, 1998.

[35] T. J. Stieltjes. Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys.*, 8(4):J1–J122, 1894.

[36] Z. Strakoš. A note on matching moments model reduction and approximation of the scattering amplitude. *An unpublished note*, 2010.

[37] Z. Strakoš and Tichý P. On efficient numerical approximation of the bilinear form $c^*a^{-1}b$. *Submitted for publication in SIAM Journal on Scientific Computing*, 2010.

[38] Z. Strakoš. Model reduction using the Vorobyev moment problem. *Numer. Algorithms*, 51(3):363–379, 2009.

[39] E. C. Titchmarsh. *Eigenfunction expansions associated with second-order differential equations. Part I.* Second Edition. Clarendon Press, Oxford, 1962.

[40] V. Totik. Orthogonal polynomials. *Surv. Approx. Theory*, 1:70–125 (electronic), 2005.

[41] Y. V. Vorobyev. *Method of moments in applied mathematics.* Gordon and Breach Science Publishers, New York, 1965. Translated from the Russian by Bernard Seckler.