University of Potsdam

Bachelor Thesis

Investigation of Functionals of the Eigenvalues of Unitary Matrices

Author:
Carina Seidel

Supervisor:
Dr. Thomas Mach

August 3, 2025

Abstract

The eigenvalues of large matrices are of interest in a large variety of use cases. However, their computation becomes increasingly challenging as matrix size grows. This thesis focuses on the investigation of functionals over the eigenvalues of unitary matrices, with particular emphasis on the spectral density, or density of states (DoS) among those. While established methods exist for real symmetric matrices, this work extends these approaches to unitary matrices with the help of the Cayley transform. The aim is to develop and analyze efficient techniques for approximating spectral functionals in this broader context.

Contents

1 Introduction **2**
- 1.1 Unitary Matrices . . . 2
- 1.2 Cayley Transform . . . 3
- 1.3 Spectral Density . . . 5

2 Regularization of the Spectral Density **7**
- 2.1 Regularization and Polynomial Expansion . . . 7
- 2.2 Orthogonal Polynomial Expansions and Basis Selection . . . 8
- 2.3 Efficient Numerical Methods . . . 9

3 Kernel Polynomial Method **11**
- 3.1 Basic Properties of Chebyshev Polynomials . . . 11
- 3.2 Approximating the Spectral Density . . . 13
- 3.3 Estimating the Trace of Chebyshev Polynomials . . . 14
- 3.4 The Kernel Polynomial Method Algorithm . . . 15
- 3.5 The KPM for Unitary Matrices . . . 15

4 Results **18**
- 4.1 Transformation of the Spectral Density for Comparison . . . 18
- 4.2 The Choice of Random Unitary Matrices . . . 19

5 Appendix **21**
- 5.1 The Jacobian for the Cayley Transform . . . 21
- 5.2 Source Code and Materials . . . 21
- 5.3 Declaration of Independent Work . . . 22
- 5.4 Zusammenfassung (Deutsch) . . . 22

Chapter 1

Introduction

The importance of eigenvalues in mathematics and physics is well established. For very large matrices, however, computing all eigenvalues can be prohibitively expensive. In many applications, it suffices to know only a few extremal eigenvalues, certain functionals of the eigenvalues, or the overall distribution of eigenvalues, leading to the concept of the spectral density.

The spectral density, as defined below, provides an exact representation of the eigenvalue distribution. Because it is formulated using the Dirac delta distribution, which is not a conventional function, it must be regularized for practical computations. While efficient techniques exist for real symmetric matrices, our aim is to extend these ideas to the unitary case. To do so, we first revisit Hermitian matrices, introduce the Cayley transform as a link to unitary matrices, and then develop the spectral density framework for unitary matrices. This foundation enables us to address the main algorithm.

1.1 Unitary Matrices

We begin with a basic definition of real symmetric matrices, which we then extend to the complex case. The superscript ^{T} denotes the transpose of a matrix, while * indicates the *conjugate transpose*, also called the *adjoint*, obtained by taking the transpose and then complex conjugating all entries. As is common in the literature, I_n denotes the identity matrix of size n.

Definition 1 (Symmetric matrix)**.** Let $A \in \mathbb{R}^{n\times n}$ be a real, square matrix of size n. Then A is called *symmetric* if $A^T = A$.

This concept can be naturally extended from real matrices to the complex case.

Definition 2 (Hermitian matrix)**.** Let $A \in \mathbb{C}^{n\times n}$ be a complex square matrix of size n. Then A is called *Hermitian* if $A^* = A$.

Throughout this thesis, A denotes a complex, square matrix of size n, unless stated otherwise. When referring to a Hermitian matrix, we use the letter H.

For an examination of the eigenvalues of a Hermitian matrix, let $H = H^* \in \mathbb{C}^{n\times n}$ and $Hv = \lambda v$ for a complex vector $v \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ and a scalar $\lambda \in \mathbb{C}$. Consider now the inner product v^*v. With $(AB)^* = B^*A^*$ and obviously $(A^*)^* = A$ we have that

$$\lambda v^* v = v^* (\lambda v) = v^* (Hv) = (v^* H) v = (H^* v)^* v = (Hv)^* v = (\lambda v)^* v = \overline{\lambda} v^* v.$$

Since we required that $v \neq \mathbf{0}$, it follows that $v^* v \neq \mathbf{0}$ and therefore we can divide the equation by the inner product to obtain $\lambda = \overline{\lambda}$, that is to say λ is real. This means that all eigenvalues of Hermitian matrices are real numbers. It follows that all eigenvalues of symmetric matrices are also real numbers, since they are a special case of Hermitian matrices.

We now introduce the types of complex matrices that play a central role in this thesis.

Definition 3 (Unitary matrix)**.** A matrix $A \in \mathbb{C}^{n \times n}$ is called *unitary* if it satisfies $A^* A = A A^* = I_n$.

The real analogue is an *orthogonal matrix*, defined by the condition $A^T A = A A^T = I_n$. We oftentimes denote unitary matrices by using U as a reference. Consider a unitary matrix U and an eigenpair (λ, v) of U. The complex conjugate of the eigenvalue equation $Uv = \lambda v$ is

$$v^* U^* = v^* \overline{\lambda} = \overline{\lambda} v^*.$$

Again, we calculate the inner product

$$v^* v = v^* I_n v = v^* U^* U v = v^* \overline{\lambda} \lambda v = \overline{\lambda} \lambda v^* v = |\lambda|^2 v^* v.$$

With the identical argument as above, we can divide by $v^* v$ to obtain

$$1 = |\lambda|^2 \implies |\lambda| = 1,$$

meaning that all eigenvalues of unitary matrices have a length of 1 and are thus situated on the unit circle. As they are a special case of unitary matrices, the same goes for orthogonal matrices. This property is crucial, as it enables the application of the Cayley transform, introduced in the following section.

There is a special group of matrices that all the matrices we have defined so far belong to.

Definition 4 (Normal matrix)**.** A matrix $A \in \mathbb{C}^{n \times n}$ is called *normal* if it commutes with its conjugate transpose, that is to say $A^* A = A A^*$.

It is straightforward to see that both Hermitian and unitary matrices are normal matrices and that the notion includes real symmetric and orthogonal matrices as special cases. The spectral theorem states that normal matrices can be diagonalized by a unitary matrix [1, Thm. 7.24, p. 218]. That means, for any normal matrix A, there exists a unitary matrix U such that $A = U \Lambda U^*$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1, \ldots, \lambda_n$ being the eigenvalues of A. This is essential for applying a function to a matrix, which is discussed in the next section.

1.2 Cayley Transform

Before giving the central definition of this section, we first extend the concept of functions to normal matrices, allowing us to map one matrix to another.

Definition 5 (Matrix function of normal matrices)**.** Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix and let $f : \mathbb{C} \to \mathbb{C}$ be a function that is defined on the spectrum of A, $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$. Then the *matrix function* $f(A)$ is defined as

$$f(A) := f(U\Lambda U^*) = U f(\Lambda) U^* = U \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n)) U^*,$$

where U is the matrix of eigenvectors of A and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues.

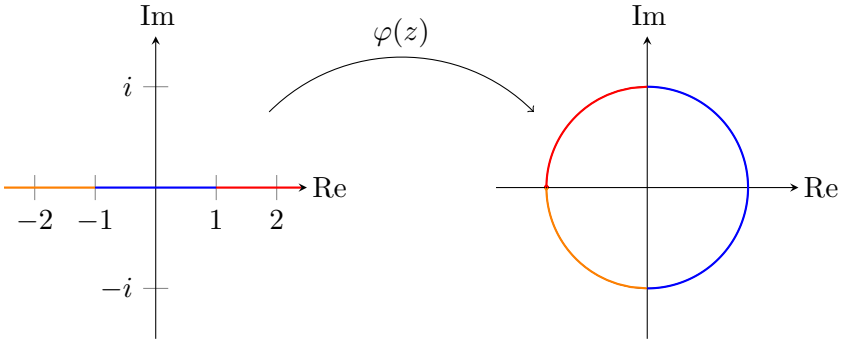
This definition is a special case of the more general definition of matrix functions via the Jordan canonical form given by Higham in [2, p. 3]. Another consequence of the spectral theorem is that for normal matrices, the Jordan canonical form is diagonal. Therefore, Higham's general definition reduces to the spectral decomposition used here.

This enables us to define the *Cayley transform*, which we then use as a matrix function that establishes a correspondence between Hermitian and unitary matrices, allowing spectral properties to be translated between these two important classes.

For a complex number $z \in \mathbb{C}$ with $z \neq -i$, the Cayley transform is defined as

$$\varphi(z) = \frac{i - z}{i + z}.$$

This function maps the real line to the unit circle in the complex plane.



The Cayley transform is a bijection from the complex plane without the point $-i$ to the unit circle, and it is continuous and differentiable everywhere except at $-i$. This specifically implies it being an involution, meaning that applying it twice returns the original value, i.e., $\varphi(\varphi(z)) = z$ for all $z \in \mathbb{C} \setminus \{-i\}$.

Let us apply the Cayley transform to Hermitian matrices. According to Definition 5, for a Hermitian matrix H with spectral decomposition $H = U\Lambda U^*$, the Cayley transform as a matrix function is given by

$$\begin{aligned}
\varphi(H) = U \operatorname{diag}(\varphi(\lambda_1), \ldots, \varphi(\lambda_n))\, U^* &= U \operatorname{diag}\left(\frac{i-\lambda_1}{i+\lambda_1}, \ldots, \frac{i-\lambda_n}{i+\lambda_n}\right) U^* \\
&= U(iI_n - \Lambda)(iI_n + \Lambda)^{-1} U^* \\
&= U(iI_n - \Lambda)U^* U(iI_n + \Lambda)^{-1} U^* \\
&= (iI_n - H)\,(iI_n + H)^{-1}
\end{aligned}$$

This calculation also illustrates that $(iI_n - H)$ and $(iI_n + H)^{-1}$ commute, or more precisely,

$$(iI_n - H)(iI_n + H)^{-1} = (iI_n + H)^{-1}(iI_n - H),$$

as they are both products of the same unitary matrix U and diagonal matrices, which all commute with each other. Consequently, the interpretation of $\frac{1}{i+z}$ (as a scalar or matrix function) does not affect the result, since the order of multiplication is immaterial in this context.

We have obtained an explicit formula which yields a unitary matrix V from a Hermitian matrix H, given that $iI_n + H$ is invertible. The Cayley transform for a Hermitian matrix is thus defined as

$$V = \varphi(H) = (iI_n - H)(iI_n + H)^{-1}.$$

The condition of $iI_n + H$ being invertible is the same as requiring that H does not have $-i$ as an eigenvalue. If (and only if) that were the case, we had $Hv = -iv$ for some eigenvector $v \neq \mathbf{0}$, and therefore

$$(iI_n + H)v = iI_n v + Hv = iv + (-iv) = 0.$$

Since Hermitian matrices only have real eigenvalues as discussed above, $-i$ can never be an eigenvalue. Conversely, given a unitary matrix $U \neq -I_n$, the inverse Cayley transform yields a Hermitian matrix H as follows:

$$H = i(I_n - U)(I_n + U)^{-1}.$$

This is relevant later, as we can then use φ to transform unitary matrices into Hermitian ones and vice versa.

The Cayley transform is not only of theoretical interest; it has been successfully applied in efficient numerical algorithms for eigenvalue problems, see for example [3] and [4].

1.3 Spectral Density

To get to the notion of the spectral density, we first need some more basic definitions to build on. The reader is assumed to be familiar with the concepts of distributions. Should that not be the case, [5] gives a comprehensive introduction in his first chapter. Let us also recap that for $\Omega \subset \mathbb{C}^n$ open and non-empty, a *test function* is a smooth function with compact support defined on Ω. The space of all test functions on Ω is usually denoted by $\mathcal{E}$.

One of the most fundamental distributions in analysis, and central to the spectral density framework, is the Dirac delta.

Definition 6 (Dirac delta distribution)**.** Let $\mathcal{E} = \mathcal{C}^\infty(\Omega)$ with $0 \in \Omega \subset \mathbb{R}^n$. Then

$$\delta : \mathcal{E} \to \mathbb{R}, \quad f \mapsto f(0) \quad \text{with} \quad \langle \delta, f \rangle := \delta(f) = f(0).$$

This distribution is often mistakenly referred to as a function, although it is not a function in the classical sense. The Dirac delta is characterized by its defining property,

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\delta(a-x)\,\mathrm{d}x = f(a) \implies \int_{-\infty}^{\infty} \delta(x-a)\,\mathrm{d}x = 1.$$

This means that the Dirac delta distribution is zero everywhere except at the point a, where it is *infinitely large*, such that the integral over it equals 1. With this, we now have all the tools we need to introduce the central concept of this thesis, the spectral density.

Definition 7 (Spectral density)**.** Let $H \in \mathbb{C}^{n\times n}$ be Hermitian. For $x \in \mathbb{R}$, the *spectral density* $\phi(x)$ is defined as

$$\phi(x) = \frac{1}{n}\sum_{j=1}^{n} \delta(x - \lambda_j),$$

where δ is the Dirac delta distribution and λ_j are the eigenvalues of H in non-descending order.

The summation over the n eigenvalues of H creates a spike at each eigenvalue λ_j, resulting in a distribution that is zero everywhere except at the eigenvalues of H. The division by n ensures that the spectral density is normalized, meaning that the Lebesgue integral over the entire real line equals 1. This is an important property, as it allows us to interpret the spectral density as a probability density function.

In practice, since $\phi(x)$ is defined using the Dirac delta distribution, it must be regularized to obtain a meaningful and computable expression. The details of this relaxation are discussed in the next chapter.

Lastly, we need the notion of the Schwartz space, which we will use as the space of test functions in the following sections.

Definition 8 (Schwartz space over $\mathbb{R}$)**.** The *Schwartz space* over $\mathbb{R}$ consists of all smooth functions f that decay rapidly to zero as $|x|$ approaches infinity [6]. Formally,

$$\mathcal{S}(\mathbb{R}) := \left\{ f \in \mathcal{C}^\infty(\mathbb{R}) \mid \forall p, k \in \mathbb{N}_0 : \sup_{x\in\mathbb{R}} \left| x^p f^{(k)}(x) \right| < \infty \right\}.$$

Chapter 2

Regularization of the Spectral Density

Because of its ability to capture the distribution of eigenvalues, the spectral density of a matrix is a crucial concept in various fields. For example, in quantum mechanics, it describes the density of energy states [7, 8]; in signal processing, it describes how power is distributed over frequency [9, 10]; and in vibration analysis, it characterizes how energy is distributed among the natural frequencies of a structure [11]. However, two main challenges arise when defining and computing the spectral density for a matrix H.

Firstly, the eigenvalues of H are typically not known in advance. If they were, computing the spectral density would be straightforward, but this is rarely the case for large matrices. Secondly, the spectral density is defined using the Dirac delta distribution, which is not a conventional function and cannot be evaluated pointwise. This makes direct computation infeasible, especially for large matrices where eigenvalue decomposition is prohibitively expensive.

2.1 Regularization and Polynomial Expansion

To address the second challenge, a fundamental idea is to replace the delta distribution by a regular function f with similar properties. To be effective, the regularizing function should be normalized, meaning it satisfies the condition

$$\int_{\mathbb{R}} f(x)\,dx = 1;$$

localized, meaning it is significant only near the origin; and sufficiently smooth to avoid introducing artificial oscillations or numerical instability. Sufficiently smooth is to say that the regularizing function should have continuous derivatives up to at least a certain order. Making sure f decays rapidly away from zero ensures that the regularized spectral density reflects the true distribution of eigenvalues without excessive broadening.

The choice of regularizing function f is crucial for both the accuracy and stability of spectral density approximations. While a single regular function f to replace the Dirac delta distribution might be hard to find, a powerful approach is to expand the spectral density in terms of a basis of functions such as $f_1, f_2, \ldots$ satisfying

the above mentioned requirements to allow for greater flexibility and accuracy. A perfect approximation would satisfy equality in the sense of distributions, meaning that it holds when integrated against a test function $g \in \mathcal{S}(\mathbb{R})$, the Schwartz space of rapidly decreasing functions from Definition 8. More precisely, this can be expressed as

$$\int_{-\infty}^{\infty} \phi(x)g(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \mu_k f_k(x)g(x)\,\mathrm{d}x.$$

If this equation holds true for all test functions g, it can be simplified to

$$\phi(x) = \sum_{k=0}^{\infty} \mu_k f_k(x), \tag{2.1}$$

which we call our *polynomial expansion*. Now we need to determine the coefficients μ_k in the expansion. The easiest way to do this is to isolate a single coefficient μ_k, as we will do in this following section.

2.2 Orthogonal Polynomial Expansions and Basis Selection

For a given interval $[a, b]$, we consider a non-negative weight function $w : [a, b] \to \mathbb{R}$. The inner product for two functions $f, g : [a, b] \to \mathbb{R}$ is then defined as

$$\langle f, g \rangle_w := \int_a^b w(x)f(x)g(x)\,\mathrm{d}x. \tag{2.2}$$

Now define $\hat{\phi}_w$ as the product of spectral density with the inverse of one such weight function w:

$$\hat{\phi}_w(x) := w^{-1}(x)\,\phi(x) = w^{-1}(x)\,\frac{1}{n}\sum_{j=1}^{n} \delta(x - \lambda_j) \tag{2.3}$$

Similarly to (2.1), we can then express the modified spectral density $\hat{\phi}_w$ as an infinite series of regular functions f_k:

$$\hat{\phi}_w(x) = \sum_{k=0}^{\infty} \mu_k f_k(x) \tag{2.4}$$

To isolate a single coefficient μ_k, it is useful that the basis functions f_k are *orthogonal* with respect to some weight function $w(x)$. This means that for any two indices k and l, we have:

$$\langle f_k, f_l \rangle_w = \int_a^b w(x)f_k(x)f_l(x)\,\mathrm{d}x = 0 \quad \text{for } k \neq l.$$

It is also advantageous to use an orthogonal polynomial base which satisfies an explicitly known three-term recurrence relation of the form

$$f_k(x) = (A_k x + B_k)f_{k-1}(x) + C_k f_{k-2}(x),$$

where $A_k, B_k, C_k \in \mathbb{R}$ are coefficients that may depend on k and the weight function. This three-term recurrence relation would allow us to efficiently compute the polynomials by using only vector instead of matrix multiplications. This entails the polynomials being degree-graded, that is to say $\deg(f_k) = k$ for each k.

It follows from Favard's theorem [12, Thm. 2.1] that for every possible choice of weight function w and interval $[a, b]$, there exists a orthogonal polynomial basis $\{f_k\}_{k=0}^{\infty}$ that is unique up to normalization and satisfies the above properties. Since there are infinitely many choices for the weight function, we have infinitely many choices for the basis functions, and we can choose one that is convenient for our purposes. Some well-studied examples of orthogonal polynomial bases are the Chebyshev polynomials, Legendre polynomials, and Hermite polynomials. These are widely used in numerical analysis and approximation theory due to their favorable convergence properties and computational efficiency [13]. We focus on Chebyshev polynomials in the next chapter.

Finally, to ensure that our chosen regularizing functions provide meaningful and accurate approximations of the spectral density in practice, we turn to practical strategies for validation and computation. These considerations are the focus of the next section.

2.3 Efficient Numerical Methods

For real symmetric matrices, a variety of efficient methods have been developed to approximate the spectral density, often relying on this regularization principle [14, 15, 16]. However, many problems in mathematics and physics naturally lead to complex Hermitian or even unitary matrices, for which the spectral density is less straightforward to compute. In this thesis, we generalize the established approaches for Chebyshev polynomials from the real symmetric case to the Hermitian case, and further broaden the scope to unitary matrices. This is achieved by first applying the Cayley transform, which provides a bridge between Hermitian and unitary matrices and allows us to transfer techniques and insights between these classes.

One intuitive approach by [15] is to select an interval $I \subset \mathbb{R}$ containing the spectrum of H, and divide it into k subintervals using points $\{t_i\}_{i=1}^{k}$. By applying Sylvester's law of inertia [17], we can count the number of eigenvalues in each subinterval, resulting in a histogram. Then we could approximate the average value of $\phi(x)$ in that interval using the fact that the number of eigenvalues in the interval can formally be expressed as

$$\nu_{[t_i,t_{i+1}]} := \int_{t_i}^{t_{i+1}} \sum_j \delta(t - \lambda_j)\, \mathrm{d}t \equiv \int_{t_i}^{t_{i+1}} n\phi(t)\, \mathrm{d}t. \tag{2.5}$$

This provides useful reference points for assessing the shape of the regularized spectral density and evaluating whether the chosen functions f_k yield an appropriate approximation.

Using Sylvester's law of inertia requires computing a decomposition of $A - t_i I = LDL^T$ for all t_i [16]. While this is efficient for certain structured matrices [18], it becomes computationally expensive or impractical for large, unstructured matrices.

This motivates the development and use of efficient numerical methods that can approximate the spectral density using only matrix-vector multiplications, which

scale better with matrix size. In this thesis, we explore one such method, the so-called Kernel polynomial method or KPM for short, which makes the computation of spectral densities feasible for large-scale problems. These advances have the potential to impact a wide range of scientific and engineering disciplines where understanding the spectral properties of large matrices is crucial.

Chapter 3

Kernel Polynomial Method

The Kernel Polynomial Method (KPM) is a powerful technique for approximating spectral properties of large matrices [14, 15]. It employs polynomial expansions to efficiently estimate quantities such as the spectral density and related functions. We previously motivated the replacement of the Dirac delta distribution with a basis of regular functions f_k. We will focus on Chebyshev polynomials for this purpose.

3.1 Basic Properties of Chebyshev Polynomials

We now develop an algorithm for the KPM that incorporates the regularization of the spectral density using Chebyshev polynomials of the first kind. As mentioned in the previous chapter, other polynomial bases could also be considered, which is why the KPM has several variants.

First, let's look how Chebyshev polynomials are defined. Using the trigonometric functions, they can be expressed as follows:

$$T_k(x) = \begin{cases} \cos(k \arccos(x)) & \text{for } k \in [-1, 1], \\ \cosh(k \operatorname{arcosh}(x)) & \text{for } k > 1, \\ (-1)^k \cosh(k \operatorname{arcosh}(-x)) & \text{for } k < -1. \end{cases}$$

For being able to make use of the polynomials' orthogonality, we only use the formula $T_k(x) = \cos(k \arccos(x))$. This means to only consider matrices that have eigenvalues within the interval $[-1, 1]$. In the case that this condition should not be fulfilled, the eigenvalues can be transformed accordingly.

For this, let λ_{lb} and λ_{ub} be the lower and upper bound for the eigenvalues of A, respectively. To find these, well-established methods like the Gershgorin circle theorem can be used [19]. Define

$$c := \frac{\lambda_{lb} + \lambda_{ub}}{2} \quad \text{and} \quad d := \frac{\lambda_{ub} - \lambda_{lb}}{2}$$

Then, the matrix $B = \frac{A - c * I_n}{d}$ has eigenvalues in the interval $[-1, 1]$. A visualization of this is linked in the bibliography.

Now, let

$$h(x) := \frac{1}{\sqrt{1 - x^2}}$$

be a weight function on $[-1, 1]$. We investigate the inner product of two Chebyshev polynomials with respect to this weight function:

Lemma 1 (Orthogonality of Chebyshev Polynomials)**.** Let $T_k(x)$ denote the Chebyshev polynomials, and let $h(x) = \frac{1}{\sqrt{1-x^2}}$ be the weight function on $[-1, 1]$. Then, for all integers $k, l \geq 0$, we have that

$$\langle T_k, T_l \rangle_h = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} T_k(x) T_l(x)\, \mathrm{d}x = \begin{cases} 0 & \text{if } k \neq l, \\ \pi & \text{if } k = l = 0, \\ \frac{\pi}{2} & \text{if } k = l \neq 0. \end{cases} \tag{3.1}$$

Proof. Recall that $T_k(x) = \cos(k \arccos(x))$ for $x \in [-1, 1]$. Let $x = \cos\theta$ with $\theta \in [0, \pi]$. Then $\mathrm{d}x = -\sin\theta\, d\theta$ and $\sqrt{1-x^2} = \sin\theta$. Substituting, we have:

$$\begin{aligned} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} T_k(x) T_l(x)\, \mathrm{d}x &= \int_{\theta=\pi}^{0} \frac{1}{\sin\theta} \cos(k\theta)\cos(l\theta)(-\sin\theta) d\theta \\ &= \int_{0}^{\pi} \cos(k\theta)\cos(l\theta) d\theta \end{aligned}$$

The orthogonality of cosine functions on $[0, \pi]$ gives:

$$\int_{0}^{\pi} \cos(k\theta)\cos(l\theta) d\theta = \begin{cases} 0 & \text{if } k \neq l, \\ \pi & \text{if } k = l = 0, \\ \frac{\pi}{2} & \text{if } k = l \neq 0. \end{cases}$$

This is a standard result in Fourier analysis [20] which proves the lemma. □

Lastly, we need a three-term recurrence relation for the Chebyshev polynomials. This is a well-known property of these polynomials, which can be derived using trigonometric identities.

Lemma 2 (Three-term recurrence for Chebyshev polynomials)**.** For all integers $k \geq 2$, the Chebyshev polynomials of the first kind satisfy

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x),$$

where the starting conditions are

$$T_0(x) = 1 \quad \text{and} \quad T_1(x) = x.$$

Proof. Let $T_k(x) = \cos(k \arccos(x))$ for $x \in [-1, 1]$ and set $\theta = \arccos(x)$, so $T_k(x) = \cos(k\theta)$ and $x = \cos\theta$. The starting conditions are straightforward to verify, as $T_0(x) = \cos(0) = 1$ and $T_1(x) = \cos(\arccos(x)) = x$. For $k \geq 2$, we use the angle addition and subtraction formulas:

$$\begin{aligned} \cos(k\theta) &= \cos((k-1)\theta + \theta) = \cos((k-1)\theta)\cos\theta - \sin((k-1)\theta)\sin\theta \\ \cos((k-2)\theta) &= \cos((k-1)\theta - \theta) = \cos((k-1)\theta)\cos\theta + \sin((k-1)\theta)\sin\theta \end{aligned}$$

Adding these two equations gives:

$$\cos(k\theta) + \cos((k-2)\theta) = 2\cos\theta\cos((k-1)\theta)$$

which can be rearranged as

$$\cos(k\theta) = 2\cos\theta\cos((k-1)\theta) - \cos((k-2)\theta)$$

Substituting back $T_k(x) = \cos(k\theta)$ and $x = \cos\theta$, we obtain

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$$

as claimed. □

3.2 Approximating the Spectral Density

Define $\hat{\phi}_h$ with the inverse of the weight function h as in (2.3), that is

$$\hat{\phi}_h(x) := h^{-1}(x)\,\phi(x) = \sqrt{1-x^2}\,\frac{1}{n}\sum_{j=1}^{n}\delta(x-\lambda_j).$$

Again, as in (2.4), we can express $\hat{\phi}_h$ as an infinite series of Chebyshev polynomials:

$$\hat{\phi}_h(x) = \sum_{k=0}^{\infty}\mu_k T_k(x)$$

Now utilize the orthogonality (3.1) of the Chebyshev polynomials to calculate a specific coefficient μ_k. Note that δ_{k0} in this context denotes the Kronecker delta, and therefore equals 1 for $k = 0$ and 0 for $k \neq 0$. Applying the inner product (2.2) to both sides of the equation yields

$$\begin{aligned}
&\left\langle \sum_{l=0}^{\infty}\mu_l T_l(x), T_k(x)\right\rangle_h = \left\langle \hat{\phi}(x), T_k(x)\right\rangle_h \\
\implies &\int_{-1}^{1}\frac{1}{\sqrt{1-x^2}}\left(\sum_{l=0}^{\infty}\mu_l T_l(x)\right)T_k(x)\,\mathrm{d}x = \int_{-1}^{1}\frac{1}{\sqrt{1-x^2}}\hat{\phi}(x)T_k(x)\,\mathrm{d}x.
\end{aligned}$$

On the left side, every term of the sum except for the k-th term vanishes due to the orthogonality of the Chebyshev polynomials, while we make use of our definition of $\hat{\phi}$ on the right side. Therefore, we have

$$\mu_k\frac{\pi}{2-\delta_{k0}} = \int_{-1}^{1}\frac{1}{\sqrt{1-x^2}}\sqrt{1-x^2}\phi(x)T_k(x)\,\mathrm{d}x \implies \mu_k = \frac{2-\delta_{k0}}{\pi}\int_{-1}^{1}\phi(x)T_k(x)\,\mathrm{d}x$$

By applying the Dirac delta distribution we obtain:

$$\begin{aligned}
\mu_k = \frac{2-\delta_{k0}}{\pi}\int_{-1}^{1}\phi(t)T_k(t)\,\mathrm{d}t &= \frac{2-\delta_{k0}}{\pi}\int_{-1}^{1}\frac{1}{n}\sum_{j=1}^{n}\delta(t-\lambda_j)T_k(t)\,\mathrm{d}t \\
&= \frac{2-\delta_{k0}}{n\pi}\sum_{j=1}^{n}T_k(\lambda_j) \\
&= \frac{2-\delta_{k0}}{n\pi}\operatorname{Tr}(T_k(A))
\end{aligned}$$

with $\mathrm{Tr}(T_k(A)) := \sum_{j=1}^{n} T_k(\lambda_j)$ the *trace* of the Chebyshev polynomial applied to the matrix A. This means we are in need of an estimator for $\mathrm{Tr}(T_k(A))$.

3.3 Estimating the Trace of Chebyshev Polynomials

We want to obtain the corollary of the following generalized theorem [21]:

Theorem 1. Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix with spectral decomposition

$$A = U\Lambda U^* \quad \text{where} \quad UU^* = I_n \text{ and } \Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$$

Let $\beta, v \in \mathbb{C}^n$ with $v = U\beta$. Suppose v is a random vector whose entries v_i are independent and identically distributed standard complex normal variables, i.e., $v_i \sim_{\text{i.i.d.}} \mathcal{N}_{\mathbb{C}}(0,1)$, meaning $\mathrm{Re}(v_i), \mathrm{Im}(v_i)$ are independent $\mathcal{N}(0, \frac{1}{2})$. Then

$$\mathbb{E}[v] = 0 \quad \text{and} \quad \mathbb{E}[vv^*] = I_n, \tag{3.2}$$

and it follows that

$$\mathbb{E}[\beta] = 0 \quad \text{and} \quad \mathbb{E}[\beta\beta^*] = I_n.$$

Proof. Since the expectation operator is linear, it holds that

$$\mathbb{E}[v] = \mathbb{E}[U\beta] = U\mathbb{E}[\beta] = 0 \implies \mathbb{E}[\beta] = 0$$

Furthermore it holds that

$$I_n = \mathbb{E}[vv^*] = \mathbb{E}[(U\beta)(U\beta)^*] = \mathbb{E}[U\beta\beta^*U^*] = U\mathbb{E}[\beta\beta^*]U^*$$

Multiplying both sides with U^* and U yields:

$$U^*I_nU = U^*U\mathbb{E}[\beta\beta^*]U^*U = \mathbb{E}[\beta\beta^*]$$

Since U is unitary, we have shown that $\mathbb{E}[\beta\beta^*] = I_n$. □

This theorem has a nice corollary when investigating a matrix function $f(A)$. In that case,

$$\begin{aligned}
\mathbb{E}\left[v^*f(A)v\right] = \mathbb{E}\left[(U\beta)^*f(U\Lambda U^*)(U\beta)\right] &= \mathbb{E}\left[\beta^*U^*Uf(\Lambda)U^*U\beta\right] \\
&= \mathbb{E}\left[\beta^*f(\Lambda)\beta\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{n} |\beta_j|^2 f(\lambda_j)\right] \\
&= \sum_{j=1}^{n} f(\lambda_j)\mathbb{E}\left[|\beta_j|^2\right] \\
&= \sum_{j=1}^{n} f(\lambda_j)
\end{aligned}$$

or, more concisely,

$$\mathbb{E}\left[v^*f(A)v\right] = \mathrm{Tr}(f(A)).$$

We now have a method to calculate the trace of a matrix function $f(A)$, using only vector multiplications with A. This is useful to estimate the trace of Chebyshev polynomials which we need to calculate our coefficients μ_k.

3.4 The Kernel Polynomial Method Algorithm

Now let $n_{\text{vec}} \in \mathbb{N}$ and consider random vectors $v_0^{(1)}, v_0^{(2)}, \ldots, v_0^{(n_{\text{vec}})}$, each drawn independently from the standard normal distribution, that is, $\mathbb{E}[v_0^{(k)}] = 0$ and $\mathbb{E}\left[v_0^{(k)} \left(v_0^{(k)}\right)^*\right] = I_n$. The subscript 0 indicates the vectors have not been multiplied by A. It follows from $\mathbb{E}\left[v^* f(A) v\right] = \operatorname{Tr}(f(A))$ that

$$\zeta_k = \frac{1}{n_{\text{vec}}} \sum_{l=1}^{n_{\text{vec}}} \left(v_0^{(l)}\right)^T T_k(A) v_0^{(l)}$$

is a good estimator for $\operatorname{Tr}(T_k(A))$ and therefore

$$\mu_k \approx \frac{2 - \delta_{k0}}{n\pi} \zeta_k.$$

In order to determine the ζ_k, let $v_0 \equiv v_0^{(l)}$. Using the recursion formula for Chebyschev polynomials, we can calculate

$$T_{k+1}(A) v_0 = 2A T_k(A) v_0 - T_{k-1}(A) v_0.$$

For $v_k \equiv T_k(A) v_0$ it also holds that

$$v_{k+1} = 2A v_k - v_{k-1}.$$

With this we are fully equipped for the final calculation and the goal of the KPM is reached: Instead of having to multiply matrices with other matrices, it now suffices to multiply matrices with vectors. Now we can approximate $\phi(x)$ as closely as we like. As aforementioned, it is not always desirable to have an infinitely exact approximation. By the Riemann-Lebesgue lemma [22, p. 103] it holds that

$$\lim_{k \to \infty} \mu_k \to 0$$

and we are only interested in $T_k(x)$ with $k \leq M$. Therefore we estimate ϕ with

$$\tilde{\phi}_M(x) := \frac{1}{\sqrt{1 - x^2}} \sum_{k=0}^{M} \mu_k T_k(x) \tag{3.3}$$

Note that this approximation no longer satisfies the condition of non-negativity, as the Chebyshev polynomials oscillate between -1 and 1. This will need to be considered when interpreting the results. Before we can state the final algorithm, we need to discuss how to apply the KPM to unitary matrices, as using the Cayley transform entails additional computations.

3.5 The KPM for Unitary Matrices

Applying the Cayley transform to an arbitrary unitary matrix U yields a Hermitian matrix H, so the KPM can in principle be applied to unitary matrices as well. While explicitly forming $H = i(I_n - U)(I_n + U)^{-1}$ works for smaller matrices, it is highly impractical for larger ones: even if U is sparse, the inverse $(I_n + U)^{-1}$ is typically

dense, making H expensive to store and destroying any sparsity. Instead, the Cayley transform is applied implicitly, as an operator acting on vectors. Whenever the algorithm requires the action of H on a vector v, we compute

$$Hv = (I_n - U)(I_n + U)^{-1}v$$

by first solving the linear system

$$(I_n + U)w = v \tag{3.4}$$

for w, and then evaluating $i(I_n - U)w$. This approach avoids explicit matrix inversion and leverages fast matrix-vector products and sparse linear solvers, preserving efficiency and scalability for large-scale problems. For large, sparse U, iterative methods such as GMRES or MINRES can be used to solve $(I_n+U)w = v$ efficiently, especially when a good preconditioner is available. This operator-based application is a key distinction from the Hermitian case, where simple matrix-vector products suffice.

We can even further reduce the number of matrix-vector products by the following observation:

$$(I_n + U)w = v \implies w + Uw = v \implies Uw = v - w.$$

So evaluating $i(I_n - U)w$ can be simplified to $i(w - Uw) = i(w - (v - w)) = i(2w - v)$. This means only having to solve the linear system once and then only using vector operations to obtain the next vector in the recursion, never even having to multiply directly with a matrix.

Another consideration that needs to be made is regarding the eigenvalues of U. Because of our using Chebyshev polynomials for orthogonality, we established that a Hermitian matrix needs to have eigenvalues in the interval $[-1, 1]$. If we were starting with a Hermitian matrix, this would be but a simple transformation. How does this translate to unitary matrices?

Algorithm 1 is based on [15, p. 10] and summarizes the steps described above. The implementation is done in Python, and linked in the appendix.

Algorithm 1 The Kernel Polynomial Method

Require: $U \in \mathbb{C}^{n\times n}$ with $U^{-1} = U^*$
Ensure: Estimated spectral density $\{\tilde{\phi}_M(t_i)\}$

for $k = 0 : M$ **do**
  $\zeta_k \leftarrow 0$
end for
for $l = 1 : n_{\text{vec}}$ **do**
  Choose a random new vector $v_0^{(l)}$; ▷ $v_{0_i}^{(l)} \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$
  for $k = 0 : M$ **do**
    Calculate $\zeta_k \leftarrow \zeta_k + \left(v_0^{(l)}\right)^T v_k(l)$;
    if $k = 0$ **then**
      Solve $(I_n + U)w_0^{(l)} = v_0^{(l)}$
      $v_1^{(l)} \leftarrow i(2w_0^{(l)} - v_0^{(l)})$
    else
      Solve $(I_n + U)w_k^{(l)} = v_k^{(l)}$
      $v_{k+1}^{(l)} \leftarrow 2i(2w_k^{(l)} - v_k^{(l)}) - v_{k-1}^{(l)}$ ▷ Three term recursion
    end if
  end for
end for
for $k = 0 : M$ **do**
  $\zeta_k \leftarrow \frac{\zeta_k}{n_{\text{vec}}}$
  $\mu_k \leftarrow \frac{2-\delta_{k0}}{n\pi}\zeta_k$
end for
Evaluate $\tilde{\phi}_M(t_i)(3.3)$

That concludes the description of the Kernel Polynomial Method. In the next chapter, we use this algorithm to generate some results, which we then compare with the actual spectral density. Furthermore, we discuss the choice of random unitary matrices, and how it influences the results.

Chapter 4

Results

We now want to use the algorithm from the previous section to compute the spectral density of a unitary matrix U. In order to assess the results, we need to compare the computed spectral density with the theoretical one. The problem is that the spectral density we defined applies to the real line, while the eigenvalues of a unitary matrix lie on the unit circle in the complex plane. To overcome this, we need to transform it a little first.

4.1 Transformation of the Spectral Density for Comparison

The exact spectral density at a point x on the real line is given by

$$\phi(x) = \sum_{i=1}^{n} \delta(x - \lambda_i),$$

where δ is the Dirac delta function. For a regularized, smoothed delta function, one simply evaluates the regularization function at $x - \lambda_i$ for all i and sums the result. We use Gaussian regularization in our implementation, with the Gaussian function defined as

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

where σ is the standard deviation of the Gaussian. This σ is often called the *target resolution* [15]; the smaller it is, the higher the resolution of the spectral density. We need to balance high resolution for exactness with easier approximation for larger σ. Then we replace the Dirac delta function in the spectral density with the Gaussian regularization function:

$$\phi_\sigma(x) := \sum_{i=1}^{n} g_\sigma(x - \lambda_i).$$

However, we defined the spectral density only on the real line, while the eigenvalues of a unitary matrix lie on the unit circle. To be able to evaluate $\phi_\sigma(z)$ for $z \in \mathbb{C}$, we use the inverse Cayley transform $\varphi^{-1}(z) = i\frac{1-z}{1+z}$, which maps $z \in \mathbb{S}^1$ to $x \in \mathbb{R}$, and evaluate $\phi_\sigma(\varphi^{-1}(z))$. According to the standard change-of-variables formula for probability densities, this requires multiplying by the Jacobian determinant to

ensure correct normalization:

$$\hat{\phi}_\sigma(z) = \phi_\sigma(\varphi^{-1}(z)) \frac{(\varphi^{-1}(z))^2 + 1}{2} = \sum_{i=1}^{n} g_\sigma(\varphi^{-1}(z) - \lambda_i) \frac{(\varphi^{-1}(z))^2 + 1}{2}.$$

An exact computation for the Jacobian determinant is found in the appendix.

Before we can plot the spectral density of a unitary matrix U on the unit circle, we need to consider which random unitary matrices to use for our numerical experiments. The choice of random unitary matrices is crucial, as it affects the distribution of eigenvalues and thus the spectral density we want to compute.

4.2 The Choice of Random Unitary Matrices

When generating random unitary matrices for numerical experiments, it is important to ensure that their eigenvalues are distributed uniformly on the unit circle, as predicted by random matrix theory [23]. A common approach is to generate a random complex matrix A with independent standard normal entries and then construct a unitary matrix from A using either the QR decomposition or the singular value decomposition (SVD). In the QR approach, A is factored as $A = QR$, where Q is unitary and R is upper triangular. The matrix Q is then used as the random unitary matrix. In the SVD approach, A is factored as $A = U\Sigma V^*$, and either U or V (both unitary) can be used as a random unitary matrix. Although both methods produce unitary matrices, their statistical properties differ. The SVD-based approach yields unitary matrices whose eigenvalues are uniformly distributed on the unit circle, matching the theoretical prediction for random unitary matrices (the so-called Haar measure) [23]. In contrast, the QR-based approach does not generally produce a uniform distribution of eigenvalues on the unit circle. This distinction is important for applications where the spectral properties of random unitary matrices play a central role, such as in the study of spectral densities.

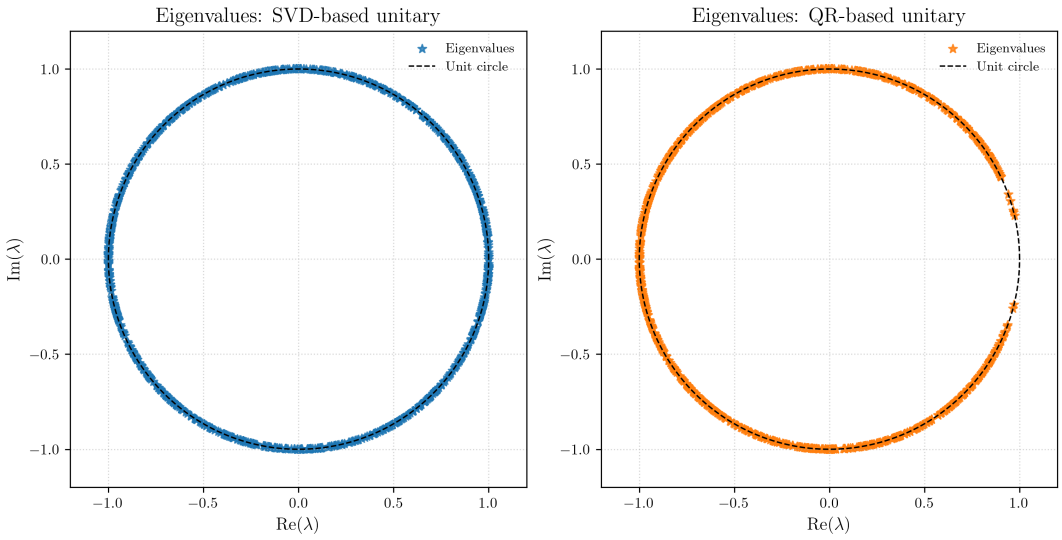


Figure 4.1: Eigenvalue distributions of random unitary matrices generated via SVD and QR.

We can now compare the spectral density of a SVD-based random unitary matrix U with that of a QR-based random unitary matrix Q. We also show how the σ influences the plots.

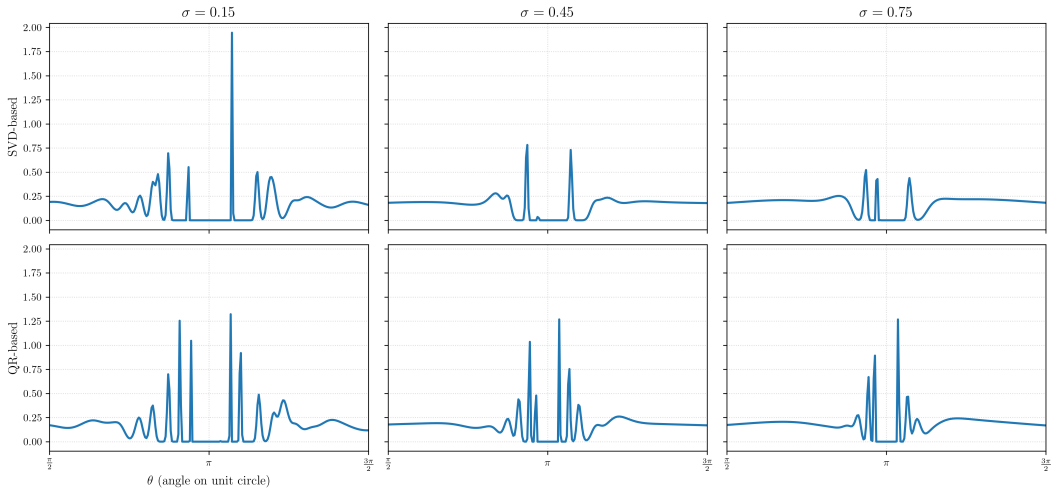


Figure 4.2: Regularized spectral density of random unitary matrices generated via SVD and QR for varying σ.

It is evident, that a bigger σ leads to a smoother spectral density but at the cost of losing information, while a smaller σ captures more fine-grained features. This behavior is consistent across both SVD-based and QR-based random unitary matrices.

Chapter 5

Appendix

5.1 The Jacobian for the Cayley Transform

To compute the Jacobian for the Cayley transform $\varphi : \mathbb{R} \to S^1$, let $x \in \mathbb{R}$. Then

$$\varphi(x) = \frac{i-x}{i+x} = -\frac{x^2-1}{x^2+1} + i\frac{2x}{x^2+1}.$$

The derivative of φ and its inverse are computed as follows:

$$\hat{\varphi} = \left[-\tfrac{x^2-1}{x^2+1}, \quad \tfrac{2}{x^2+1}\right] \implies \frac{\mathrm{d}\hat{\varphi}}{\mathrm{d}x} = \left[-\tfrac{4x}{(x^2+1)^2}, \quad \tfrac{2-2x^2}{(x^2+1)^2}\right]$$

by the quotient rule. The norm of the derivative is then

$$\left\| \frac{\mathrm{d}\hat{\varphi}}{\mathrm{d}x} \right\| = \sqrt{\left(-\frac{4x}{(x^2+1)^2}\right)^2 + \left(\frac{2-2x^2}{(x^2+1)^2}\right)^2} = \sqrt{\frac{16x^2+4-8x^2+4x^4}{(x^2+1)^4}} = \frac{2}{x^2+1}.$$

Therefore, the Jacobian for the transformation from z to w is

$$J(z) = \left|\frac{dw}{dz}\right| = \frac{2}{(z^2+1)}$$

and the inverse Jacobian is $\frac{z^2+1}{2}$.

Thus, to transform the spectral density from the real line to the unit circle, we multiply by

$$\frac{x^2+1}{2}$$

This factor accounts for the change of variables and ensures that the density is properly normalized on the unit circle. (See handwritten notes for the derivation.)

5.2 Source Code and Materials

Below is the link to our GitHub repository, which contains both the LaTeX source files for this thesis and the executable, self-written Python code implementing the Kernel Polynomial Method.
`https://github.com/ClarinaSaxel/spectral-density-of-unitary-matrices`

5.3 Declaration of Independent Work

I hereby declare that I have written this thesis independently and have not used any sources or aids other than those indicated. All passages taken from other works, either verbatim or in substance, have been identified as such.

5.4 Zusammenfassung (Deutsch)

Diese Bachelorarbeit behandelt die numerische Approximation der Spektraldichte unitärer Matrizen mit der Kernel-Polynom-Methode (KPM). Es werden die mathematischen Grundlagen, die Regularisierung der Spektraldichte, die Anwendung der KPM sowie numerische Ergebnisse vorgestellt und diskutiert. Die Arbeit wurde in englischer Sprache verfasst; diese Zusammenfassung erfolgt gemäß den formalen Anforderungen in deutscher Sprache.

Bibliography

[1] S. Axler, *Linear Algebra Done Right*, Springer, 2015.

[2] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, 2008.

[3] J. L. Laurent, T. Mach, R. Vandebril, D. S. Watkins, *Computing the eigenvalues of symmetric tridiagonal matrices via a Cayley transformation*, Electronic Transactions on Numerical Analysis, 46: 447–459, 2017.

[4] K. Meerbergen, A. Spence, D. Roose, *Shift-invert and Cayley transforms for detection of rightmost eigenvalues of nonsymmetric matrices*, Numerische Mathematik, 67(2): 161–189, 1994.

[5] R. S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms*, World Scientific, 2003.

[6] R. D. Richtmyer, *Principles of Advanced Mathematical Physics*, vol 1, Springer Verlag, New York, 1981.

[7] L. E. Reichl, *A Modern Course in Statistical Physics*, 4th edition, Wiley-VCH, 2016.

[8] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, 2nd edition, Addison-Wesley, 2011.

[9] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd edition, Prentice Hall, 2010.

[10] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 2nd edition, McGraw-Hill Science Engineering Math, 2001.

[11] M. P. Norton, D. G. Karczub, *Fundamentals of Noise and Vibration Analysis for Engineers*, Cambridge University Press, 2003.

[12] G. Szegő, *Orthogonal Polynomials*, 4th edition, American Mathematical Society, Colloquium Publications, Vol. 23, 1975.

[13] J. C. Mason and D. C. Handscomb, *Chebyshev Polynomials*, Chapman & Hall/CRC, 2003.

[14] A. Weisse, G. Wellein, A. Alvermann, H. Fehske, *The kernel polynomial method*, Reviews of Modern Physics, 78(1): 275–306, 2006.

[15] L. Lin, Y. Saad, C. Yang, *Approximating Spectral Densities of Large Matrices*, SIAM Review, 58(1): 34–65, 2016.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th edition, Johns Hopkins University Press, Baltimore, MD, 2013.

[17] J. J. Sylvester, *A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares*, Philosophical Magazine, 4(23): 138–142, 1852.

[18] P. Benner and T. Mach, *Computing all or some eigenvalues of symmetric $\mathcal{H}_l$-matrices*, SIAM Journal on Matrix Analysis and Applications, 33(4): 1395–1415, 2012.

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd edition, Cambridge University Press, 2013.

[20] E. Kreyszig, *Advanced Engineering Mathematics*, 10th edition, John Wiley & Sons, Hoboken, NJ, 2011.

[21] M. F. Hutchinson, *A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines*, Communications in Statistics - Simulation and Computation, 18(3): 1059–1076, 1989.

[22] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, 1987.

[23] F. Mezzadri, *How to generate random matrices from the classical compact groups*, Notices of the AMS, 54(5), 592–604, 2007.

[24] G. W. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2010.

[25] L. Pastur, M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, AMS, 2011.

[26] C. Seidel, *Annäherung an die Spektraldichte großer Matrizen*, unpublished, Universität Potsdam, 2024.