University of Potsdam

Bachelor Thesis

Investigation of Functionals of the Eigenvalues of Unitary Matrices

Author:
Carina Seidel

Supervisor:
Dr. Thomas Mach

July 14, 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

Introduction

We begin this thesis by examining unitary matrices and their fundamental properties. Then we revisit the Cayley transform to finally close in on the spectral density. Upon this we should be well-equipped to proceed with the investigation afterwards.

1.1 Unitary Matrices

We first recall two definitions for important real matrices that we then extend to complex matrices. The index ^{T} marks the transpose of a matrix. As common in literature, I_n denotes the identity matrix of size n.

Definition 1 (Symmetric matrix)**.** Let A be a real, square matrix of size n. Then A is called *symmetric* if $A^T = A$.

The following definition is for matrices with their transpose as their inverse.

Definition 2 (Orthogonal matrix)**.** Let A be a real, square matrix of size n. Then A is called *orthogonal* if $A^T \cdot A = I_n$.

The complex equivalent of a real symmetric matrix is a *Hermitian matrix.* Note that A^* is *conjugate transpose* of the matrix A with all of its entries complex conjugated and transposed.

Definition 3 (Hermitian matrix)**.** Let A be a complex square matrix of size n. Then A is called *Hermitian* if $A^* = A$.

Throughout this thesis, A will denote a complex, square matrix of size n, unless stated otherwise. To reference a Hermitian matrix, we will use the letter H.

We examine the eigenvalues of Hermitian matrices.

Let $H = H^*$ and $Hv = \lambda v$ for a complex vector $v \neq \mathbf{0}$ of size n and a scalar $\lambda \in \mathbb{C}$. Consider now the inner product v^*v.

$$\lambda v^* v = v^* (\lambda v) = v^* (Hv) = (v^* H) v = (H^* v)^* v = (Hv)^* v = (\lambda v)^* v = \overline{\lambda} v^* v \quad (1.1)$$

Since we have that $v \neq \mathbf{0}$ it follows that $v^*v \neq \mathbf{0}$ and therefore $\lambda = \overline{\lambda}$, 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. Now we can extend the definition of orthogonal matrices to complex matrices.

Definition 4 (Unitary matrix)**.** A matrix A is called *unitary* if $A^* \cdot A = I_n$.

We will oftentimes denote unitary matrices by using U as a reference. It is easy to see that orthogonal matrices are a special case of unitary matrices, since $A^T = A^*$ for all real matrices.
Consider a unitary matrix U and an eigenpair (λ, v) of U. The complex conjugate of the eigenvalue equation $Uv = \lambda v$ is

$$v^* U^* = v^* \overline{\lambda} = \overline{\lambda} v^* \tag{1.2}$$

We calculate

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

Similarly to above, we can divide by $v^* v$ to obtain

$$1 = |\lambda|^2 = |\lambda| \tag{1.3}$$

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.

1.2 Cayley Transform

Before deep diving into the exact definition of the Cayley transform, we will first introduce the concept of a *matrix function.*

Definition 5 (matrix function)**.** Let A be a real square matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then the *matrix function* $f(A)$ is defined as

$$f(A) = U f(\Lambda) U^T$$

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

The Cayley transform establishes a correspondence between Hermitian and unitary matrices, allowing spectral properties to be translated between these two important classes.

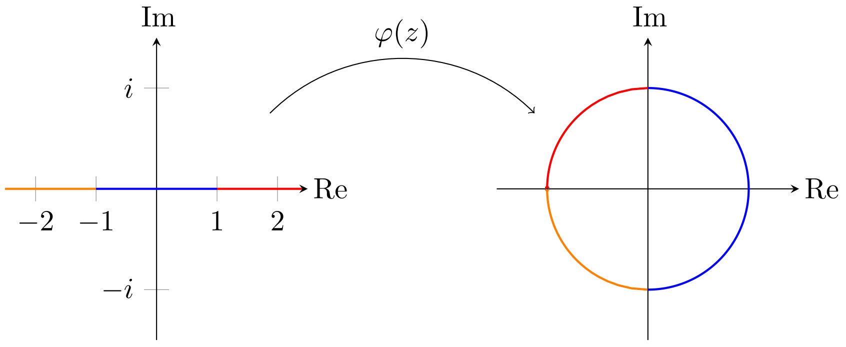
For a complex number $z \in \mathbb{C}$ with $z \neq -i$, the Cayley transform is defined as

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

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

For matrices, the Cayley transform maps a Hermitian matrix H (with $i + H$ invertible) to a unitary matrix U via

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



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

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

Luckily, we know that Hermitian matrices can only ever have real eigenvalues, so $-i$ can never be an eigenvalue.

Conversely, given a unitary matrix U (with $U \neq -I_n$), the inverse Cayley transform yields a Hermitian matrix:

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

This will be relevant later, as we can then use φ to transform unitary matrices into symmetric ones and vice versa. In particular, it enables the transfer of spectral density results, which will be explored in the following sections.

1.3 Spectral Density

To get to the notion of the spectral density, we will first need some more basic definitions to build on.

Definition 6 (linear functional)**.** Let V be a vectorspace over a field $\mathbb{K}$. A *linear functional* T is a linear function $T : V \to \mathbb{K}$. The space over all linear functionals $V \mapsto \mathbb{K}$ is called the *dual space* V'.

A simple example for such a functional would be

$$T : \mathcal{C}^\infty(\mathbb{R}) \to \mathbb{R}, \qquad f \mapsto f(0) \tag{1.4}$$

A more special case is the integral.

$$T_g : \mathcal{C}^\infty(\mathbb{C}) \to \mathbb{C}, \qquad f \mapsto \int\limits_{\mathbb{C}} g \cdot f \, \mathrm{d}x \tag{1.5}$$

This definition leads to the concept of a distribution which was introduced to get a method to differentiate where differentiation in the classical sense is not possible.

Definition 7 (distribution)**.** Let $\emptyset \neq \Omega \subset \mathbb{R}^n$ be open. Let $\mathcal{E}$ be the space of *test functions* over Ω. A *distribution* T is a function $T : \mathcal{E} \to \mathbb{C}$ where for all $g, g_1, g_2, \{g_n\}_{n\in\mathbb{N}} \in \mathcal{E}$
with $\lim\limits_{n\to\infty} g_n \to g$ it holds:

$$T(g_1 + \lambda g_2) = T(g_1) + \lambda T(g_2) \quad \text{and} \quad \lim_{n\to\infty} T(g_n) \to T(g)$$

In summary, a distribution T is a continuous and linear functional on $\mathcal{E}$. This leads us to the definition of the Dirac delta distribution:

Definition 8 (Dirac delta function)**.** Let $\mathcal{E} = \mathcal{C}^\infty(\Omega)$ with $0 \in \Omega \subset \mathbb{R}^n$. Then

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

An important property of the Dirac delta is:

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

This distribution is often mistakenly referred to as a function, although it is not a function in the classical sense.
We can now define the spectral density:

Definition 9 (Spectral density)**.** Let $A \in \mathbb{R}^{n\times n}$, $A^T = A$, and A sparse. The spectral density is defined as

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

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

The number of eigenvalues in an interval $[a, b]$ can then be expressed as:

$$\nu_{[a,b]} = \int\limits_{a}^{b} \sum_{j} \delta(t - \lambda_j)\,\mathrm{d}t \equiv \int\limits_{a}^{b} n\phi(t)\,\mathrm{d}t \tag{1.6}$$

Definition 10 (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 [4]. Formally,

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

Motivation

2.1 Motivation

Calculating the spectral density of a matrix is straightforward when its eigenvalues are already known. However, this is rarely the case in practice, and computing the eigenvalues of very large matrices is both time-consuming and computationally expensive. At the same time, the density of states (DOS), which serves as a probability density for the distribution of eigenvalues, is of great interest in many fields. Therefore, there is a need for efficient methods to approximate the spectral density at low computational cost.

The challenge arises because $\phi(t)$, the delta distribution, is not a conventional *function* that can be evaluated pointwise.

One intuitive approach is to select an interval $I \in \mathbb{R}$ such that the spectrum of A, $\sigma(A)$, is contained within I. Next, choose k points t_i in I so that the interval is divided into subintervals:

$$\{t_i\}_{i=1}^{k} \subset I \quad \text{with} \quad \bigcup_{i=1}^{k-1} [t_i, t_{i+1}] = I$$

Count the number of eigenvalues in each subinterval. Then, calculate the average value of $\phi(t)$ in each interval using $\nu_{[a,b]}$ from equation 1.6. The result is a histogram, which, as the subintervals become smaller (i.e., as k increases and $(t_{i+1} - t_i) \longrightarrow 0$), approaches the true spectral density.

To count the eigenvalues in the intervals, one can use methods such as Sylvester's law of inertia. The details of this method are beyond the scope of this work; it would require computing a decomposition of $A - t_i I = LDL^T$ for all t_i [3]. Instead, we prefer a method in which A is multiplied by vectors, which is more efficient in higher dimensions.

Kernel Polynomial Method

3.1 Overview

The so called kernel polynomial method, or KPM for short, has many variants and is a powerful tool for approximating the spectral density of matrices. We will focus on the main approach in the following.
As the name suggests, the KPM is a polynomial extension of the spectral density. The coefficients of the polynomials are derived from the method of moments, in order to obtain an estimator function as in statistics. The method is based on a corollary of the following theorem:

Theorem 1. Let $A = A^T \in \mathbb{R}^{n\times n}$ with spectral decomposition

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

Also, let $\beta, v \in \mathbb{R}^n$ with $v = U\beta$.
If $v_i \sim_{\text{i.i.d.}} \mathcal{N}(0,1)$ for the components $\{v_i\}_{i=1}^n$ of v, that is to say

$$\mathbb{E}[v] = 0 \text{ and } \mathbb{E}[vv^T] = I_n, \tag{3.1}$$

then

$$\mathbb{E}[\beta] = 0 \text{ and } E[\beta\beta^T] = I_n$$

Proof of Theorem 1. It holds that

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

Furthermore it holds that

$$\mathbb{E}[vv^T] = \mathbb{E}[U\beta\beta^T U^T] = U\mathbb{E}[\beta\beta^T]U^T = U\mathbb{E}[\beta\beta^T]U^T = I_n$$

Since U is orthogonal, we can multiply both sides with U^T and U:

$$U^T\mathbb{E}[vv^T]U = \mathbb{E}[\beta\beta^T] = U^T I_n U = I_n$$

Thus, we have shown that $\mathbb{E}[\beta\beta^T] = I_n$. □

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

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

also zusammengefasst

$$\mathbb{E}\left[v^T f(A) v\right] = \operatorname{Spur}(f(A)) \tag{3.2}$$

3.2 Polynomiale Erweiterung durch Tschebyschev-Polynome

Aufgrund ihrer vielen einzigartigen Eigenschaften sind Tschebyschev-Polynome besonders gut zur polynomialen Erweiterung der Delta-Distribution geeignet. Mit Hilfe der trigonometrischen Funktionen können sie auch wie folgt ausgedrückt werden:

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

Wir benutzen im Folgenden nur die Formel $T_k(t) = \cos(k \arccos(t))$. Daher müssen wir uns auf Matrizen beschränken, deren Eigenwerte im Intervall $[-1, 1]$ liegen. Sollte diese Voraussetzung nicht erfüllt sein, kann man die Eigenwerte entsprechend transformieren. Seien dazu λ_{us} und λ_{os} jeweils die untere bzw. obere Schranke für die Eigenwerte von A. Definiere

$$c := \frac{\lambda_{us} + \lambda_{os}}{2} \quad \text{und} \quad d := \frac{\lambda_{os} - \lambda_{us}}{2}$$

Dann ist $B = \frac{A - c * \mathbb{1}_n}{d}$ eine Matrix mit Eigenwerten im Intervall $[-1, 1]$. Eine Veranschaulichung dazu ist im Anhang verlinkt.

Tschebyschev-Polynome können zudem mit der Rekursionsformel

$$T_{k+1}(t) = 2t T_k(t) - T_{k-1}(t)$$

berechnet werden, wobei die Startbedingunen $T_0(t) = 1$ und $T_1(t) = x$ gelten.

Beachte auch, dass das Resultat in Gleichung 3.2 besagt, dass

$$\mathbb{E}\left[v^T T_k(A) v\right] = \sum_{j=1}^{n} T_k(\lambda_j) = \operatorname{Spur}(T_k(A)) \tag{3.3}$$

gilt. Dies ist zentral im weiteren Vorgehen.

Sei nun

$$h(x) = \frac{1}{\sqrt{1-t^2}} \tag{3.4}$$

eine Gewichtsfunktion. Eine weitere Eigenschaft der Tschebyschev-Polynome ist, dass sie *orthogonal* bezüglich des mit h gewichteten Skalarproduktes

$$\langle f, g \rangle = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} \cdot f(x) \cdot g(x) \, \mathrm{d}x$$

sind. Das bedeutet, dass

$$\int_{-1}^{1} \frac{1}{\sqrt{1-t^2}} \cdot T_k(t) \cdot T_l(t) \, \mathrm{d}t = \begin{cases} 0 & \text{für } k \neq l \\ \pi & \text{für } k = l = 0 \\ \frac{\pi}{2} & \text{für } k = l \neq 0 \end{cases}$$

3.3 Annäherung der Spektraldichte

Multipliziere nun die Spektraldichte mit dem Inversen der Gewichtsfunktion 3.4:

$$\hat{\phi}(t) = \sqrt{1-t^2}\phi(t) = \sqrt{1-t^2} \times \frac{1}{n} \sum_{j=1}^{n} \delta(t - \lambda_j)$$

Sei nun $g \in \mathcal{S}$, dem in Definition 10 beschriebenen Schwartz-Raum, und $\mu_k \in \mathbb{R}$ Koeffizienten, die wir nachher berechnen, sodass die folgende Gleichung gilt:

$$\int\limits_{-1}^{1} \hat{\phi}(t) g(t) \, \mathrm{d}t = \int\limits_{-1}^{1} \sum_{k=0}^{\infty} \mu_k T_k(t) g(t) \, \mathrm{d}t \tag{3.5}$$

Gilt dies für beliebige $g \in \mathcal{S}$, so vereinfachen wir Gleichung 3.5 zu

$$\hat{\phi}(t) = \sum_{k=0}^{\infty} \mu_k T_k(t) \tag{3.6}$$

Nutze nun die Orthogonalität der Tschebyschev-Polynome aus, um einen bestimmten Koeffizienten μ_k zu berechnen:

$$\begin{aligned} \sum_{l=0}^{\infty} \mu_l T_l(t) = \hat{\phi}(t) \implies & \left(\sum_{l=0}^{\infty} \mu_l T_l(t) \right) \cdot T_k(t) = \hat{\phi}(t) \cdot T_k(t) \\ \implies & \int_{-1}^{1} \frac{1}{\sqrt{1-t^2}} \cdot \left(\sum_{l=0}^{\infty} \mu_l T_l(t) \right) \cdot T_k(t) \, \mathrm{d}t = \int_{-1}^{1} \frac{1}{\sqrt{1-t^2}} \cdot \hat{\phi}(t) \cdot T_k(t) \, \mathrm{d}t \\ \implies & \mu_k \cdot \frac{\pi}{2 - \delta_{k0}} = \int_{-1}^{1} \frac{1}{\sqrt{1-t^2}} \cdot \sqrt{1-t^2} \cdot \phi(t) \cdot T_k(t) \, \mathrm{d}t \\ \implies & \mu_k = \frac{2 - \delta_{k0}}{\pi} \cdot \int_{-1}^{1} \phi(t) \cdot T_k(t) \, \mathrm{d}t \end{aligned}$$

Durch Anwendung der Delta-Funktion erhält man:

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

Sei nun $n_{vec} \in \mathbb{R}$ und $v_0^{(1)}, v_0^{(2)}, \ldots, v_0^{(n_{vec})}$ Vektoren, die die Bedingungen aus dem Theorem erfüllen, also $\mathbb{E}[v_0^{(k)}] = 0$ und $\mathbb{E}\left[v_0^{(k)} \left(v_0^{(k)}\right)^T\right] = \mathbb{1}_n$. Aus Gleichung 3.3 folgt, dass

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

ein guter Schätzer für $\operatorname{Spur}(T_k(A))$ ist und damit

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

Um die ζ_k zu bestimmen, sei im Folgenden $v_0 \equiv v_0^{(l)}$ Berechne nun mit Hilfe der Rekursionsformel für Tschebyschev-Polynome:

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

Für $v_k \equiv T_k(A)v_0$ gilt also, dass

$$v_{k+1} = 2Av_k - v_{k-1}$$

Damit sind alle Bauteile zur Berechnung festgelegt und das Ziel der KPM erreicht: Anstatt rechenaufwendig Matrizen mit anderen Matrizen zu multiplizieren, müssen wir sie nur noch mit Vektoren multiplizieren. Nun können wir $\phi(t)$ beliebig nah annähren. Wie bereits erwähnt, ist eine unendlich genaue Annäherung nicht immer wünschenswert. Wegengilt

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

und wir interessieren uns nur für $T_k(t)$ mit $k \leq M$
Daher schätzen wir ϕ durch

$$\tilde{\phi}_M(t) = \frac{1}{\sqrt{1-t^2}} \sum_{k=0}^{M} \mu_k T_k(t) \tag{3.7}$$

Der folgende Pseudocode basiert auf [2, p. 10] und fasst die oben beschriebenen Schritte zusammen. Ich habe ihn selbst implementiert und im Anhang verlinkt.

Algorithm 1 Die Kernel-Polynom-Methode

```
Require: $A = A^T \in \mathbb{R}^{n \times n}$ mit Eigenwerten aus dem Intervall $[-1, 1]$
Ensure: Geschätzte Spektraldichte $\{\tilde{\phi}_M(t_i)\}$

    for $k = 0 : M$ do
        $\zeta_k \leftarrow 0$
    end for
    for $l = 1 : n_{\text{vec}}$ do
        Wähle einen neuen zufälligen Vektor $v_0^{(l)}$;          ▷ $v_{0_i}^{(l)} \sim_{\text{i.i.d.}} \mathcal{N}(0, 1)$
        for $k = 0 : M$ do
            Berechne $\zeta_k \leftarrow \zeta_k + \left(v_0^{(l)}\right)^T v_k(l)$;
            if $k = 0$ then
                $v_1^{(l)} \leftarrow A v_0^{(l)}$
            else
                $v_{k+1}^{(l)} \leftarrow 2 A v_k^{(l)} - v_{k-1}^{(l)}$          ▷ Drei-Term-Rekursion
            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
    Werte $\tilde{\phi}_M(t_i)$ mit Gleichung 3.7 aus
```

Bibliography

[1] Christian Bär, *Lineare Algebra und analytische Geometrie*, Springer Spektrum 2018

[2] Lin Lin, Youssef Saad and Chao Yang, *Approximating Spectral Densities of Large Matrices*, `https://arxiv.org/pdf/1308.5467v2`
5. Oktober 2014

[3] G. H. Golub und C. F. Van Loan, *Matrix Computations, 4th Edition*, Johns Hopkins University Press, Baltimore, MD, 3rd ed., 2013

[4] R. D. Richtmyer, *Principles of advanced mathematical physics*, vol 1, Springer Verlag, New York, 1981