BSc Thesis Applied Mathematics

Von Neumann's Inequality
A proof based on Linear Algebra

Lavinia Suvi Lanting

Supervisor: F. Schwenninger

July 1, 2021

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Preface

One of the greatest struggles of mathematics is that, while it finds many applications in a wide variety of fields, communicating it consistently proves to be rather difficult. It comes as no surprise that mathematics is a fairly unpopular subject that many find difficult to approach. What those many seem to misunderstand is that mathematics is, as a matter of fact, so broad and varied and can be approached in so many different way that it can also be reprocessed to be accessible for everyone due to its pliability. An example of this are the works by prof.dr. Nelly Litvak, who has written the book 'Who needs mathematics?' in which she approaches mathematics and its applications in digital technologies for the general public. In this book as well as in her Facebook group, Nelly Litvak approaches mathematics through the eyes of non-mathematically-minded people in order to elaborate and explain mathematics in terms accessible to them. Her effort in wanting to clearly communicate mathematics to non-mathematically-minded adults poses the question whether or not mathematics is currently being sufficiently well communicated to those outside of circle of mathematicians. More often than not, it would indeed seem as though mathematics, which is a universal language, forgets that not everybody is well versed in the higher registry and that most people posses but a limited comprehension of its complexity.

It is no secret that physics is one of the bigger fields to which mathematics is applied. Many results find use in that discipline. Such a result would be von Neumann's Inequality, which finds an application in quantum mechanics permitting a rigorous description of the theories within the field. While many physicist do have a broad training in mathematics, the question remains whether or not results utilised in the discipline are properly communicated. We may ask ourselves, as mathematicians, why we choose to communicate mathematics in a certain fashion rather than another.

In this thesis we provide a different approach to the proof of von Neumann's Inequality, aiming to give a proof which solely rests on Linear Algebra, a branch of mathematics which is taught in the first year of study and accessible even for engineering students and, therefore, more accessible than Functional Analysis, the standard tool utilised when approaching von Neumann's inequality. In doing so, we shall first approach this different proof step by step by ensuring all utilised results are based on Linear Algebra. We will then take a closer look at certain aspects of the classical proof and we will then compare the two different approaches.

Von Neumann Inequality
a proof based on Linear Algebra

Lavinia. S. Lanting*

July 1, 2021

Abstract

Von Neumann's inequality asserts that, given a contraction T operator on a Hilbert space H, the following inequality holds for any polynomial p with complex coefficients:

$$\|p(T)\| \leq \sup_{|z|\leq 1} |p(z)|. \tag{1}$$

The formulation of this theorem only requires elementary notions, yet the proof is usually approached through Functional Analysis.
In this thesis this result is tackled utilising solely Linear Algebra, including a proof of the Maximum Modulus Principle. In addition, variations on this inequality will also be touched upon.

Keywords: von Neumann, Maximum modulus principle, contraction, spectral theorem

*Email: l.s.lanting@student.utwente.nl

Contents

1 Introduction

Von Neumann's Inequality is a mathematical result belonging to Functional Analysis and Operator Theory and due to John von Neumann, a Hungarian-American mathematician, physicist, computer scientist, engineer and polymath whose aim was to integrate pure and applied sciences and who made major contributions to a myriad of fields. He was the very first to establish a rigorous mathematical framework for quantum mechanics. This framework is known as the Dirac–von Neumann axioms and was first introduced in his work *"Mathematical Foundations of Quantum Mechanics"* in 1932. While this work was more of a summary of results that von Neumann had established in earlier papers, it brought his ideas and discoveries together and reduced the physics of quantum mechanics to the mathematics of Hilbert spaces and linear operators acting on them.
Among the results present in this thesis, we also encounter von Neumann's Inequality, an inequality providing an upper bound for the norm of a polynomial of a contraction T acting on a Hilbert space H, as follows:

$$\|p(T)\| \leq \sup_{z\in\overline{\mathbb{D}}} |p(z)|, \tag{2}$$

where $\overline{\mathbb{D}}$ is the closed unit disk in the complex plane.
This result in particular will be at the centre of this thesis. Historically, this inequality has been proved by means of Functional Analysis by using the dilation of a contraction. A dilation is an operator acting on a larger Hilbert space of which H can be seen as a subspace. Roughly speaking, the operator T equals the projection of the dilation on H. As it turns out, it can be proven that every contraction has a unitary dilation, i.e. a dilation that is unitary. The validity of von Neumann's inequality for the unitary dilation of the contraction T is easily proved thanks to the unitary nature of this operator, as it follows directly from the Spectral Theorem. Just a few extra steps are then required to prove that the inequality also must hold for T itself. The structure of this proof enables it to be generally applicable to all contraction acting on a Hilbert space, independently of whether or not they are normal or unitary. As a matter of fact, von Neumann's inequality can be shown to hold for normal operators as a consequence of the Spectral Theorem, completely bypassing this proof technique. This is generally not the case, which is why in this thesis we look at the more general case of von Neumann's inequality for any contraction. In doing so, we will also approach the proof from a different angle than the historical one by solely relying on Linear Algebra rather than Functional Analysis and will then show and compare the differences between the two proofs. This would seem a natural course of action, as the statement of von Neumann's inequality only involves basic concepts from Linear Algebra. To this avail, we will first state notions and results needed to prove von Neumann's inequality and we will prove those results when needed. We will then set out to prove von Neumann's inequality and will later give insight into certain details of its historical proof and how it differs from the main result of this paper.

2 Notions and prerequisites

In the following we introduce the required concepts and results from Linear Algebra and Analysis. To this avail, it may then be of use to explicit and look back onto some of the definitions needed as basis for the upcoming results.

The formulation of von Neumann's Inequality requires us to be familiar with the concept of *bounded operator*, *contraction* and objects such as the *closed unit disk*. We recall that the closed unit disk around the point 0 is the set of points whose distance from 0 is less than or equal to one: $\overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$. In addition, we define the unit circle as follows: $\partial\mathbb{D} := \{z \in \mathbb{C} : |z| = 1\}$.

Definition 2.1 (Bounded operator). *A bounded linear operator T is a linear transformation $T : X \to Y$ that maps bounded subsets of X to bounded subsets of Y, where X and Y are normed vector spaces, vector spaces equipped with a norm. T is bounded if and only if there exists some $M \geq 0$ such that*

$$\|Tx\|_Y \leq M\|x\|_X$$

for all $x \in X$.
The smallest such M is the operator norm of T, $\|T\|$.

In this thesis, all mentioned operators are assumed to be linear bounded operators.

Definition 2.2 (Contraction). *A bounded linear operator $T : X \to Y$ between normed vector spaces X and Y is said to be a contraction if its operator norm $\|T\|$, the induced norm $\|T\| = \sup\limits_{x \neq 0} \frac{\|Tx\|}{\|x\|}$, is less or equal to 1.*

Of importance is that the von Neumann's Inequality is specifically defined for a contraction acting on a Hilbert space, the definition of which is hereby included.

Definition 2.3 (Hilbert space). *A Hilbert space H is a real or complex inner product space and complete metric space with respect to the distance function induced by the inner product.*

We recall that by a complete metric space we mean that every Cauchy sequence $x_n \in H$ converges to a point in H.
We here make notice of what the implications of this are: when talking about a vector norm we namely will always connote the square root of the *inner product*, i.e. $\|x\| = \sqrt{(x, x)}$, of the considered Hilbert space and when talking about an operator or matrix norm we will always refer to the operator norm with respect to this vector norm. Furthermore, the inner product of elements x and y is denoted as (x, y).

Definition 2.4 (Spectral Norm for Matrices). *The Spectral Matrix Norm for a matrix $A \in \mathbb{C}^{n \times m}$ is defined as the largest singular value of A, which is the square root of the largest eigenvalue of the matrix A^*A, where A^* is the complex conjugate of A. Hence*

$$\|A\| := \sqrt{\lambda_{max}(A^*A)} = \sigma_{max}(A).$$

We have here introduced the notation $\sigma_{max}(A)$, equivalent to $\sqrt{\lambda_{max}(A^*A)}$, also known as the *spectral radius* of the matrix A. For completion, we, therefore, also include the definition of the *spectrum* of a matrix.

Definition 2.5 (Spectrum of a matrix)**.** *Let V be a finite-dimensional vector space over the field $\mathbb{C}$ and suppose that $T : V \to V$ is a linear operator. The spectrum of T, $\sigma(T)$, is the set of roots of the characteristic polynomial of T. Hence, the elements of $\sigma(T)$ are the eigenvalues of T. In addition, if the matrix T has eigenvalues λ_i $(i = 1, 2, ..., n)$, $\sigma_{max}(T)$ denotes the spectral radius of T and $\sigma_{max}(A) = \max_{i=1,...,n} |\lambda_i|$.*

As it turns out, as an addition to Definition 2.4, the following equality also holds:

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|.$$

Lemma 1. *The Spectral Matrix Norm for a matrix $A \in \mathbb{C}^{n \times m}$ equals the induced norm of such that*

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|.$$

Proof. In order to prove this result, we must prove that there exists a unit vector x such that $\|A\| = \|Ax\|$ and that $\|A\| = \sup_{\|x\|=1} \|Ax\|$. It suffices here to make use of the induced matrix norm, according to which $\|A\| = \sup_{y \neq 0} \frac{\|Ay\|}{\|y\|}$. Let $x = \frac{y}{\|y\|}$, then clearly $\|x\| = 1$. Since $\|y\|$ is a scalar, it holds that $\frac{\|Ay\|}{\|y\|} = \|\frac{Ay}{\|y\|}\|$. By this we may then obtain that $\|A\| = \sup_{y \neq 0} \frac{\|Ay\|}{\|y\|} = \sup_{y \neq 0} \|\frac{Ay}{\|y\|}\| = \sup_{\|x\|=1} \|Ax\|$. □

In addition to the already mentioned definitions, we will make use of particular properties of *unitary* and *hermitian* matrices.

Definition 2.6 (Unitary, Hermitian and positive semidefinite matrices)**.** *A Unitary matrix, a Hermitian matrix and a positive semidefinite matrix are defined as follows:*

- ***Unitary matrix**: A complex square matrix $U \in \mathbb{C}^{n \times n}$ is unitary if its conjugate transpose U^* is also its inverse. That is $UU^* = U^*U = I_n$, where I_n is the identity matrix of dimensions $n \times n$.*
- ***Hermitian matrix**: A Hermitian or self-adjoint matrix is a square matrix in $\mathbb{C}^{n \times n}$ that is equal to its conjugate transpose. This means that a matrix A is Hermitian $\iff$ $A = A^*$ $\iff$ $a_{ij} = \overline{a_{ji}}$.*
- ***positive semidefinite matrices**: An $n \times n$ Hermitian matrix A is said to be positive semidefinite if and only if $\overline{x}^T Ax \geq 0$ for all $x \in \mathbb{C}^n$.*

For completion, we shall also include the definition of a unitary operator.

Definition 2.7 (Unitary operator)**.** *An operator acting on an Hilbert space is said to be unitary if its conjugate transpose U^* equals its inverse. That is $UU^* = U^*U = I$, where I is the identity operator.*

Relating to unitary operators, we also include the following two lemmas.

Lemma 2. *Let U be a unitary operator. Then U is a bijective operator. Furthermore $\|Ux\| = \|x\|$.*

Proof. The first part is a simple consequence of the operator U being unitary. We will, therefore, only take a look at the second part.
Let U be a unitary operator on a Hilbert space H and let x be in H. Then $\|Ux\|^2 = |(Ux, Ux)|$. Since U is unitary, $U^*U = I$ and $U^* = U^{-1}$, hence

$$\|Ux\|^2 = |(Ux, Ux)| = |(x, U^*Ux)| = |(x, x)| = \|x\|^2.$$

This proves the given statement. □

Finally, we include for completion the definition of an separable metric space, which will later return when tackling von Neumann's Inequality applied to an infinite dimensional Hilbert space.

Definition 2.8 (Separable metric space)**.** *A metric space X is said to be separable if and only if it contains a countable dense subset. In other words, a metric space X is separable if and only if there is a countable set Z of X such that for every point $a \in X$ there is a sequence $(x_k)_{k\in\mathbb{N}}$ in Z such that $x_k \to a$ as $k \to \infty$.*

Aside from the already mentioned definitions and lemmas, we shall also include two theorems needed later to prove von Neumann's inequality.

Theorem 3 (Polar decomposition)**.** *Let $A \in \mathbb{C}^{n\times n}$, then there exist matrices P and U, where U is a unitary matrix and P is a positive semidefinite Hermitian matrix, such that A may be factorised as follows: $A = UP$.*

Proof. Let $A \in \mathbb{C}^{n\times n}$. Let $x_1, x_2, ..., x_n$ be an orthonormal basis of eigenvectors for $\sqrt{A^*A}$, which is a positive semidefinite matrix which exists because A^*A is normal. This means that $\sqrt{A^*A}x_i = Px_i = \lambda_i x_i$, where $1 \leq i \leq n$. To demonstrate this we consider the orthonormal set of vectors $\{\frac{1}{\lambda_1}Ax_1, \frac{1}{\lambda_2}Ax_2, ..., \frac{1}{\lambda_n}Ax_n\}$, which is an orthonormal set.
We now turn this orthonormal set into the columns of a matrix and multiply it by E, the adjoined of a matrix containing our original orthonormal basis of eigenvectors for $\sqrt{A^*A}$. This yields our definition of U:

$$\begin{aligned} U &= \begin{pmatrix} \frac{1}{\lambda_1}Ax_1 & \frac{1}{\lambda_2}Ax_2 & \cdots & \frac{1}{\lambda_n}Ax_n \end{pmatrix} \cdot \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^* \\ &= \begin{pmatrix} \frac{1}{\lambda_1}Ax_1 & \frac{1}{\lambda_2}Ax_2 & \cdots & \frac{1}{\lambda_n}Ax_n \end{pmatrix} \cdot E. \end{aligned}$$

Let us define the standard unit vector $s_i \in \mathbb{C}^n$ as $[s_i]_j = 0$ whenever $i \neq j$ and $[s_i]_j = 1$ whenever $i = j$. We then investigate the matrix vector product of E with any element of the orthonormal basis for P.

$$\begin{aligned} Ex_i &= \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^* x_i \\ &= x_1^*[x_i]_1 + x_2^*[x_i]_2 + ... + x_i^*[x_i]_i + ... + x_n^*[x_i]_n \\ &= 0 + 0 + ... + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} + ... + 0 = s_i. \end{aligned}$$

This means that $Ux_i = \frac{1}{\lambda_i} Ax_i$. We now introduce P between our orthonormal matrix U and our basis eigenvector x_i. This yields:

$$UPx_i = U\lambda_i x_i = \lambda_i U x_i = \lambda_i \frac{1}{\lambda_i} Ax_i = Ax_i.$$

Clearly then $A = UP$ for the basis of eigenvectors $x_1, x_2, ..., x_n$, [2] □

The following theorem can be found in Friedman's book on principles and techniques of Applied Mathematics, [6].

Theorem 4 (Spectral Theorem)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a self-adjoint matrix. Then there exists a unitary matrix $Q \in \mathbb{C}^{n\times n}$ and eigenvalues $\lambda_1, ...\lambda_n$ of A such that $Q^*Q = I$ and $Q^*AQ = diag(\lambda_1, ...\lambda_n)$. The columns of Q form an orthonormal basis for $\mathbb{C}^n$.*

This theorem holds more generally for normal matrices, matrices which commute with their conjugate transpose.
In addition, $\|A\| = \|diag(\lambda_1, ...\lambda_n)\|$ and $\|diag(\lambda_1, ...\lambda_n)\| = \max\limits_{1\le i\le n} |\lambda_i|$ as we will prove below.

Lemma 5 (Norm of a diagonal matrix)**.** *Let D be a diagonal matrix with diagonal entries λ_i, $i = 1, ..., n$. Then for some unit vector x*

$$\|D\| = \|diag(\lambda_1, ...\lambda_n)\| = \max_{1\le i\le n} |\lambda_i|.$$

Proof. As $D = diag(\lambda_1, ...\lambda_n)$, the following holds:

$$\begin{aligned} \|D\| = \|diag(\lambda_1, ...\lambda_n)\| &= \max_{\|x\|=1} \sqrt{\sum_{i=1}^{n} \lambda_i^2 x_i^2} \\ &\le \max_{\|x\|=1} \max_{1\le i\le n} |\lambda_i| \sqrt{\sum_{i=1}^{n} x_i^2} \\ &= \max_{1\le i\le n} |\lambda_i| \max_{\|x\|=1} \|x\| \\ &= \max_{1\le i\le n} |\lambda_i|. \end{aligned}$$

Let now x be an eigenvector of D corresponding to the largest eigenvalue. Consequently $\|Ax\| = \max\limits_{1\le i\le n} |\lambda_i|$. This then yields that $\|A\| = \max\limits_{1\le i\le n} |\lambda_i|$. □

3 Maximum Modulus Principle

The previously shown definitions and results are not the only elements needed to construct the proof of the von Neumann's Inequality. The *Maximum Modulus Principle* also plays a central role in the proof. As we have set as our goal to prove the von Neumann's Inequality solely relying on Linear Algebra, it is only natural that all theorems we utilise during our proof also should be proved by Linear Algebra or must, at the very least, be provable by means of Linear Algebra solely. In this section we then set out to prove the *Maximum Modulus Principle* in such a fashion. Before tackling the *Maximum Modulus Principle* and its proof, we must mention a result we will use to demonstrate this theorem.

Lemma 6. *A square matrix U is unitary if and only if it is unitarily equivalent to a diagonal matrix D, with diagonal elements having all modulus one. In other words, U is unitary if and only if there exists a unitary matrix T such that $D = T^*UT$, where D is a diagonal matrix with diagonal elements all of modulus one. This implies that $\|U\| = \|D\| = 1$.*

Proof. This is a consequence of the Spectral Theorem, as U is normal.
Let U be unitary, by Theorem 4 there exists a unitary matrix T such that $D = T^*UT$, meaning that U and D are unitarily equivalent, since $\|U\| = \|D\| = 1$ and since $\|D\| = \max_i |d_i|$ as by Lemma 5, where d_i $(i = 1, ..., n,)$ are the diagonal entries of D. Since the diagonal elements of D are the eigenvalues of U, the statement follows since eigenvalues of a unitary matrix have to have modulus one.
On the other hand, let now U be unitarily equivalent to to a diagonal matrix D, such that the diagonal elements of D all have modulus one. Then there exists a unitary matrix T such that $D = T^*UT$. Since $\|D\| = \|T^*UT\| = \|U\|$ and since $\|D\| = 1$, it must be that $\|U\| = 1$ and that U is unitary, seeing as U is a composition of unitary matrices. □

Given the previous lemma, we are now ready to state and demonstrate the *Maximum Modulus Principle.* The statement as well as the proof will be based on Orr Moshe Shalit's article *"A sneaky proof of the maximum modulus principle"*, [11].

Theorem 7 (Maximum Modulus Principle)**.** *Let f be a polynomial analytic in a neighbourhood of the closed unit disc $\overline{\mathbb{D}}$. Then*

$$\max_{z\in\overline{\mathbb{D}}} |f(z)| = \max_{z\in\partial\mathbb{D}} |f(z)|.$$

Proof. Suppose Theorem 7 indeed holds for polynomials, then let a function f be analytic in a neighbourhood of $\overline{\mathbb{D}}$ and choose $\epsilon > 0$. There exists a polynomial p such that $\sup_{\overline{\mathbb{D}}} |f - p| < \epsilon$. Then

$$\begin{aligned}
\max_{\overline{\mathbb{D}}} |f| &= \max_{\overline{\mathbb{D}}} |p + f - p| \\
&\leq \max_{\overline{\mathbb{D}}} |p| + \max_{\overline{\mathbb{D}}} |f - p| \leq \max_{\partial\mathbb{D}} |p| + \epsilon \\
&= \max_{\partial\mathbb{D}} |f + p - f| + \epsilon \leq \max_{\partial\mathbb{D}} |f| + \max_{\partial\mathbb{D}} |p - f| + \epsilon \\
&\leq \max_{\partial\mathbb{D}} |f| + 2\epsilon.
\end{aligned}$$

For ϵ arbitrarily small it then follows that

$$\max_{\overline{\mathbb{D}}} |f| \leq \max_{\partial\mathbb{D}} |f|. \tag{3}$$

Let now f be a polynomial and let n be the degree of the polynomial f and let z be any point belonging to $\overline{\mathbb{D}}$. Set $s = \sqrt{1-|z|^2}$ and define the unitary $(n+1)\times(n+1)$ matrix

$$U = \begin{pmatrix} z & 0 & \cdots & 0 & s \\ s & 0 & \cdots & 0 & -\overline{z} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

The matrix U is unitary as

$$\begin{aligned} UU^* &= \begin{pmatrix} z & 0 & \cdots & 0 & s \\ s & 0 & \cdots & 0 & -\overline{z} \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z & s & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & \vdots & \ddots & 1 \\ s & -\overline{z} & 0 & \cdots & 0 \end{pmatrix} \\ &= \begin{pmatrix} z^2+s^2 & zs-\overline{z}s & 0 & \cdots & 0 \\ zs-\overline{z}s & s^2+\overline{z}^2 & 0 & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} z^2+1-|z|^2 & 0 & 0 & \cdots & 0 \\ 0 & 1-|z|^2+|z|^2 & 0 & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} = U^*U. \end{aligned}$$

Let now P denote the $(n+1)$ column vector of the form

$$P = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

We may then see that the equality $z^k = P^T U^k P$ holds for all $k = 1, ..., n$. Since the degree of f is n, we utilise this equality to express $f(z)$ in terms of $f(U)$ as follows:

$$f(z) = P^T f(U) P = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} f(z) & f(0) & \cdots & f(0) & f(s) \\ f(s) & f(0) & \cdots & f(0) & f(-\overline{z}) \\ f(0) & f(1) & f(0) & \cdots & f(0) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ f(0) & \cdots & f(0) & f(1) & f(0) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

$f(z) = P^T f(U) P$. Note now that

$$\|P\| = \left\| \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\| = \sqrt{1^2 + 0^2 + \ldots + 0^2} = \left\| \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \right\| = \|P^T\| = 1,$$

which yields the following result:

$$|f(z)| = |P^T f(U) P| \leq \|P^T\| \|f(U)\| \|P\| = \|f(U)\|.$$

By Lemma 6 we know that the matrix U is unitarily equivalent to a diagonal matrix $diag(w_1, ..., w_{n+1})$, where $|w_i| = 1$ for all $i = 1, ..., n+1$. Furthermore

$$f(diag(w_1, ..., w_{n+1}) = diag(f(w_1), ..., f(w_{n+1})).$$

Hence

$$|f(z)| \leq \|f(U)\| = \|diag(f(w_1), ..., f(w_{n+1}))\| = \max_{1 \leq i \leq n+1} |f(w_i)| \leq \max_{\partial \mathbb{D}} |f|.$$

Having proven that $|f(z)| \leq \max_{\partial \mathbb{D}} |f|$, combined with Equation (3) that $\max_{\overline{\mathbb{D}}} |f| \leq \max_{\partial \mathbb{D}} |f|$, indeed proves Theorem 7. $\square$

4 Von Neumann's Inequality

Now that the Maximum Modulus Principle has been proven by means of Linear Algebra, we are ready to state and prove the main result of this paper: von Neumann's Inequality. In order to prove this result, we need several lemmas, which we will state and prove hereafter.

Lemma 8. *For every matrix $A \in \mathbb{C}^{nn}$ there exist unit vectors x and y such that $\|A\| = |(Ax, y)|$ and such that $\|A\| = \sup_{|x|=1, |y|=1} |(Ax, y)|$.*

Proof. By Lemma 1 there exists a sequence (x_n) bounded in $\mathbb{C}^n$ such that $|x_n| = 1$. Therefore, there exists a subsequence and $\|Ax_n\| \to \|A\|$ as $n \to \infty$. Since $\{y \in \mathbb{C}^n : \|y\| = 1\}$ is complex, there exists a subsequence (x_{n_k}) of (x_n), which converges to x. Since A is continuous, Ax_{n_k} converges to Ax and thus $\|Ax_n\|$ converges to $\|A\|$, as A is compact.
To see the second identity, note that by Cauchy-Schwartz the following inequality holds: $\|A\| \geq \sup_{\|x\|=1, \|y\|=1} |(Ax, y)|$.
On the other hand

$$\begin{aligned} \sup_{\|x\|=1, \|y\|=1} |(Ax, y)| &\leq \sup_{\|x\|=1, \|y\|=1} \|Ax\| \|y\| \\ &= \sup_{\|x\|=1, \|y\|=1} \|Ax\| \\ &\leq \sup_{\|x\|=1, \|y\|=1} |\|A\| \|x\| = \|A\|, \end{aligned}$$

meaning that $\sup_{\|x\|=1, \|y\|=1} |(Ax, y)| \leq \|A\|$.
Since $\|A\| \geq \sup_{\|x\|=1, \|y\|=1} |(Ax, y)|$ and $\sup_{\|x\|=1, \|y\|=1} |(Ax, y)| \leq \|A\|$, the following must hold: $\sup_{\|x\|=1, \|y\|=1} |(Ax, y)| = \|A\|$.
This proves the lemma. □

We may now state and prove von Neumann's Inequality for contractions T acting on a finite dimensional Hilbert space, meaning we prove this result for the case in which the operator is represented by a matrix. The following theorems and proofs are all based on Gilles Pisier's book (Ch.1, p.13-14), [9].

Theorem 9 (von Neumann's Inequality for a finite dimensional Hilbert Space)**.** *Let T be a contraction acting on a finite dimensional Hilbert space and let p be any polynomial with complex coefficients. Then:*

$$\|p(T)\| \leq \sup_{z \in \overline{\mathbb{D}}} |p(z)|. \tag{4}$$

Proof. Let T be a contraction acting on a finite dimensional Hilbert space, meaning that the operator T maps the Hilbert space on which it is acting to itself. The operator T can, therefore, be represented by a complex square $n \times n$ matrix, i.e.

$T \in \mathbb{C}^{n\times n}$, where n equals the dimension of the space. Since T is a square matrix, we may factorise it according to the polar decomposition, Theorem 3, such that

$$T = U\overline{T},$$

where $U \in \mathbb{C}^{n\times n}$ is a unitary matrix and $\overline{T}$ is a positive semidefinite Hermitian matrix. We note that by the Spectral Theorem, Theorem 4, $\overline{T}$ is unitarily diagonalisable, since it is Hermitian, meaning that we may find a unitary matrix P such that $\overline{T} = P^*DP$, where D is a diagonal matrix. We note that

$$\|T\| = \|U\overline{T}\| = \|\overline{T}\| = \|P^*DP\| = \|P^*\|\|D\|\|P\| = \|D\|,$$

since matrices U and P (and thus also of P^*) are all unitary, meaning they have no influence whatsoever on the norm of the matrices they compose as follows from Lemma 2. Following this we obtain that $\|D\| = \|T\| \leq 1$, since T is a contraction by assumption. This implies that there exist λ_j, $j = 1, ..., n$, such that $D = diag(\lambda_1, ..., \lambda_n)$ and

$$||D|| = \left\| \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix} \right\| = \max_{1\leq j\leq n} |\lambda_j| \leq 1 \tag{5}$$

by Lemma 5. Therefore, $|\lambda_j| \leq 1$, $1 \leq j \leq n$. Let now

$$T(z_1, ..., z_n) = UP^* \begin{pmatrix} z_1 & 0 & 0 & \cdots & 0 \\ 0 & z_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & z_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & z_n \end{pmatrix} P$$

for $|z_i| \leq 1$, $i = 1, ..., n$, such that $T(\lambda_1, ..., \lambda_n) = T$. Let now $z = z_1$ and fix z_i, $i = 2, ..., n$. Then

$$T(z, z_2, ..., z_n) = UP^* \begin{pmatrix} z & 0 & 0 & \cdots & 0 \\ 0 & z_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & z_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & z_n \end{pmatrix} P$$

Now take p to be any polynomial in one variable with complex coefficients. Then $p(T(z, z_2, ..., z_n)) = (p(z))_{ij}$, $i, j = 1, 2, ..., n$.
By Lemma 8 there exist unit vectors x and y such that $\|p(T(z))\| = |(p(T(z))x, y)|$, where x and y depend on z, p and T. We note that $(p(T(z))x, y)$ is a polynomial in the variable z. Therefore, the Maximum Modulus Principle, Theorem 7, can be applied for the running variable z. This yields

$$\|p(T(z_1, ..., z_n))\| = |(p(T(z))x, y)| \leq \max_{z\in\partial\mathbb{D}} |(p(T(z))x, y)| \leq \max_{z\in\partial\mathbb{D}} \|p(T(z)\|.$$

This is done iteratively, each time fixing a different $z = z_i$ for $i = 1, ..., n$, which gives, in particular,

$$\|p(T(\lambda_1, ..., \lambda_n))\| \leq \sup_{z_1,...,z_n \in \partial\mathbb{D}} \|p(T(z_1, ..., z_n)\|.$$

We note that, for $z_1, ..., z_n \in \partial\mathbb{D}$, $(T(z_1, ..., z_n)$ is unitary by Lemma 6 and may then be diagonalised in such a way that

$$T(z_1, ..., z_n) = V^* \begin{pmatrix} z_1 & 0 & 0 & \cdots & 0 \\ 0 & z_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & z_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & z_n \end{pmatrix} V,$$

where V, and therefore also V^*, is a unitary matrix. We apply p once again to this equality and obtain that

$$p(T(z_1, ..., z_n)) = V^* \begin{pmatrix} p(z_1) & 0 & 0 & \cdots & 0 \\ 0 & p(z_2) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & p(z_{n-1}) & 0 \\ 0 & 0 & \cdots & 0 & p(z_n) \end{pmatrix} V.$$

Hence

$$\begin{aligned} \|p(T(z_1, ..., z_n))\| &= \left\| V^* \begin{pmatrix} p(z_1) & 0 & 0 & \cdots & 0 \\ 0 & p(z_2) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & p(z_{n-1}) & 0 \\ 0 & 0 & \cdots & 0 & p(z_n) \end{pmatrix} V \right\| \\ &= \left\| \begin{pmatrix} p(z_1) & 0 & 0 & \cdots & 0 \\ 0 & p(z_2) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & p(z_{n-1}) & 0 \\ 0 & 0 & \cdots & 0 & p(z_n) \end{pmatrix} \right\| \leq \sup_{z \in \partial\mathbb{D}} |p(z)|. \end{aligned}$$

Hence this proves that

$$\|p(T(\lambda_1, ..., \lambda_n))\| \leq \sup_{z \in \partial\mathbb{D}} |p(z)|.$$

In addition, by the maximum principle it is given that $\sup_{z \in \overline{\mathbb{D}}} |p(z)| = \sup_{z \in \partial\mathbb{D}} |p(z)|$. Therefore,

$$||p(T)|| = ||p(T(\lambda_1, ..., \lambda_n))|| \leq \sup_{z \in \partial\mathbb{D}} |p(z)| = \sup_{z \in \overline{\mathbb{D}}} |p(z)|,$$

which concludes the proof of our statement. □

We here make note of the fact that von Neumann's Inequality also holds for polynomials p in multiple variables, analytic in each variable. The proof follows then directly from the proof here above: instead of fixing z, we consider all z_1, $i = 1, ..., n$ and $|z_i| = 1$, at the same time. The Maximum Modulus Principle is then applied repeatedly in each variable.

Having proven von Neumann's Inequality for the finite dimensional case, we will now state and approach the proof of the infinite dimensional case, the inequality as classically stated and the result we truly aim to prove. We first though will state and prove a lemma that we will need in order to prove this more general result. To this avail, we will also include the definition of an orthogonal projection operator.

Definition 4.1 (Orthogonal projection operator)**.** *A complex operator P is an orthogonal projection if it is a linear operator such that $P^2 = P$ and $P^* = P$.*

Lemma 10. *Let P be an orthogonal projection operator. Then then norm of P is smaller or equal to one, i.e. $\|P\| \leq 1$.*

Proof. Let P be an orthogonal projection operator and x be any vector. Then

$$\|Px\|^2 \leq \|Px\|^2 + \|(I - P)x\|^2 = \|x\|^2,$$

which proves our statement as $\|P\| \leq 1$. □

Theorem 11 (Von Neumann's Inequality)**.** *Let T be a contraction acting on a Hilbert space H and let p be any polynomial with complex coefficients. Then:*

$$\|p(T)\| \leq \sup_{z \in \overline{\mathbb{D}}} |p(z)|, \tag{6}$$

where $\overline{\mathbb{D}}$ is the closed unit disk.

Proof. If H is finite dimensional, then the statement follows from Theorem 9. For simplicity of notation, we shall assume the infinite dimensional Hilbert space H to be separable. This is allowed seeing as H admits a countable orthonormal basis. Let then $\{E_n\}$ be an increasing sequence of finite dimensional subspaces of H such that $\overline{\bigcup E_n} = H$. Now let P_n be the orthogonal projection from H onto E_n and set $T_n = P_n T J_n$, where T_n is the mapping $T_n : E_n \to E_n$ and J_n is the embedding from E_n to H. This projection exists, since E_n is finite dimensional, meaning that there exists an orthonormal basis for E_n. As follows from Lemma 10, the norm of an orthogonal projection operator is smaller or equal to 1, because of which the following inequality holds:

$$\|T_n\| = \|P_n T J_n\| \leq \|P_n\| \|T\| \|J_n\| = \|T\|.$$

Hence $\|T_n\| \leq \|T\|$. By Theorem 9, we already know that $||p(T_n)|| \leq \sup_{z \in \overline{\mathbb{D}}} |p(z)|$ for any polynomial with complex coefficients in one variable. We observe now that $T_n x \to Tx$ whenever $n \to \infty$ for any $x \in H$. It then follows that $p(T_n)x \to p(T)x$ as $n \to \infty$ for all $x \in H$. From this we deduce that the following inequality holds:

$$\|p(T)\| \leq \sup_{z \in \overline{\mathbb{D}}} |p(z)|.$$

This then concludes the proof for the infinite dimensional case. □

At this point, the question might naturally arise whether or not a variation on von Neumann's Inequality exists which holds for operators of norm greater than 1. As it turns out, that is indeed the case.

Theorem 12 (Von Neumann's Inequality for any operator)**.** *Let T be any operator acting on an infinite dimensional Hilbert space and let p be any polynomial with complex coefficients, then:*

$$\|p(T)\| \leq \sup_{|z| \leq \|T\|} |p(z)|. \tag{7}$$

Proof. Consider the operator $\overline{T} = \frac{T}{\|T\|}$. Clearly $\|\overline{T}\| = \frac{\|T\|}{\|T\|} = 1$, hence $\overline{T}$ is a contraction acting on a Hilbert space H. The statement follows then from Theorem 11. □

5 Maximum Modulus Principle: a different approach

As we had previously mentioned, we wanted to consider the proof of von Neumann's Inequality to be purely based on Linear Algebra if all results utilised in the proof also are proved relying solely on Linear Algebra. In regards to Theorem 9, the proof of this theorem barely required us to utilise results belonging to Functional Analysis. It did though require us to prove certain results again in order for those not to rely anymore on Functional Analysis. As for Theorem 11, one might argue that the proof of this statement is not entirely based on Linear Algebra, as we did make use of separable metric spaces, which are concepts belonging to Analysis. As it turns out though, it is not just the proof of our main result that we have had to modify to fit our goal, but also the proof of one of the theorems utilised to approach it. The validity of Theorem 7 is usually demonstrated in a very different fashion by utilising properties belonging to Functional Analysis.
Before moving forward and analysing the two proofs side by side, we see it fit to take a short look at how the Maximum Modulus Principle is usually tackled. To this avail, we shall include and prove a lemma needed for the proof and also propose and prove the Maximum Modulus Principle as done in Edward B. Saff and Arthur David Snider's book, [10].
We must briefly recall that a domain is an open connected set in $\mathbb{C}$.

Lemma 13. *Suppose that f is analytic in a disk centred at z_0 and that the maximum value of $|f(z)|$ over this disk is $|f(z_0)|$. Then $|f(z)|$ is constant in the disk.*

Proof. Assume that $|f(z)|$ is not constant. Then there must exist a point z_1 inside the disk such that $|f(z_0)| > |f(z_1)|$. Let C_R denote the circle centred at z_0 which passes through z_1 with radius $R = |z_1 - z_0|$. Then $|f(z_0)| \geq |f(z)|$ for all z on C_R by assumption. In addition, by the continuity of f, the strict inequality $|f(z_0)| > |f(z)|$ must hold for z on a subset of C_R containing z_1. Seeing as C_R may be parametrised as $z = z_0 + Re^{it}$, this leads to a contradiction of the following equation, derived by the Cauchy formula for f:

$$\begin{aligned} f(z_0) &= \frac{1}{2\pi i} \oint_{C_R} \frac{f(z)}{z - z_0} dt \\ &= \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(z_0 + Re^{it})}{Re^{it}} iRe^{it} dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + Re^{it}) dt. \end{aligned}$$

This is because, since $|f(z_0)| > |f(z_1)|$, the subset containing z_1, when evaluated through this integral, would produce a deficit, which could only be made up by another point z_2 on C_R for which $|f(z_0)| < |f(z_2)|$. That is a clear contradiction, since z_0 is the maximum attained by the function f. Must must, therefore, be constant inside this disk. $\square$

We may now show the more classical approach to the Maximum Modulus Principle.

Theorem 14 (Maximum Modulus Principle)**.** *If f is analytic in a domain D and $|f(z)|$ achieves its maximum value at a point z_0 in D, then f is constant in D.*

Proof. Suppose that $|f(z)|$ is not constant. Then there must exist a point z_1 in D such that $|f(z_1)| < |f(z_0)|$. Let γ be a path in D running from z_0 to z_1 and consider the values of $|f(z)|$ for z on γ, starting at z_0. We should expect to find a point w along γ where $|f(z)|$ first starts to decrease. That is, there should be a point w on γ which fulfils the following properties:

- $|f(z)| = |f(z_0)|$ for all z preceding w on γ.
- There are points z on γ, arbitrarily close to w, where $|f(z)| < |f(z_0)|$.

Note that w may coincide with z_0. In addition, from the first mentioned property and the continuity of f, we have that $|f(w)| = |f(z_0)|$.
Since every point of a domain is an interior point, there must be a disk centred at w that lies in D. Since Lemma 13 applies and says that $|f|$ is constant in this disk, contradicting property the second mentioned property, we are forced to conclude that the initial supposition about the existence of z_1 must be incorrect. Consequently $|f|$, and therefore f itself, is constant in D. □

One may immediately notice that, while the are both named the Maximum Modulus Principle, Theorem 7 and Theorem 14 are extremely different. As a matter of fact, they are two 'different' theorems. Theorem 7 is a weaker version than Theorem 14, as Theorem 7 does show that the maximum of a function f, analytic on a certain domain in the shape of a disk, must be obtained on the boundary of this disk, but it mentions nothing about the behaviour of the function in the case that the maximum is obtain on the interior. Theorem 14 makes mention of this and proves that no non-constant function can reach it's maximum on the interior of said disk. What's more, the proofs are also clearly very different and not just in what sort of mathematical theory they rely upon. Theorem 7 has a rather lengthy proof that requires multiple steps, while Theorem 14 relies much more on intuition and requires few simple steps.
What is especially interesting about these differences is that, while Theorem 14 demands more advanced mathematics to be known, its proof is much easier then the one of Theorem 7. This might very well be the reason why this version of the Maximum Modulus Principle, relying on Functional Analysis, is usually utilised when proving von Neumann's inequality instead of the version presented in Theorem 14, in spite of the fact that the latter relies on Linear Algebra, a wider spread and better known branch of mathematics.

We must of course also make note of the differences present in the proof itself. The historical proof makes use of the dilation of the contraction T acting on the Hilbert space H in question. This dilation is an operator acting upon a larger Hilbert space of which H is a subspace and it can be orthogonally projected onto H in order to obtain T. Since T is a contraction, this dilation must be a unitary operator for which the validity of von Neumann's inequality follows directly from the Spectral Theorem, Theorem 4. Proving the same statement for T becomes then a direct

consequence of the previous steps.
Both methods try to accomplish the same: starting with the contraction T, one tries to find a unitary matrix relating to T for which von Neumann's inequality holds and by which its validity also can be proven for T. The main difference lies then in how this unitary matrix is found. In the paper we accomplished that by a combination of the Polar Decomposition, Theorem 3, and the Spectral Theorem, Theorem 4, while in the historical proof only one step is needed: it suffices to prove that the dilation of T is indeed unitary. The latter, while it rests on more difficult mathematics, is a simpler and shorter approach, while the result provided in this paper is certainly accessible to a much broader audience. This might very well be a reason why this proof is preferable to the historical one.

6 Variations on von Neumann's Inequality

Von Neumann's inequality shows that a contraction T acting upon an Hilbert space H is polynomially bounded. By this we mean that

$$\|p(T)\| \leq \sup_{z \in \overline{\mathbb{D}}} |p(z)|, \tag{8}$$

for any polynomial p with complex coefficients and where $\overline{\mathbb{D}}$ is the closed unit disk. One may wonder whether or not the closed unit disk is the only set which can provide such a bound and, in fact, that is not the case. Several more inequalities and bounds have been found since, even for operators that are not contractions. In this section we will first look into a possible new set Ω and will try to define its characteristics and how these relate to the other inequalities and bounds that have been defined since von Neumann's inequality. In particular, we will discuss Crouzeix's Conjecture and will reflect upon whether or not this conjecture provides a better or worse bound for an operator T acting on a Hilbert space.

6.1 About the eigenvalues of T

Firstly we are interested in knowing how this new set Ω relates to the spectrum of the operator T. By Theorem 9, we are aware that Ω must contain all eigenvalues of T whenever T is a contraction and $\Omega = \overline{\mathbb{D}}$, but can that be said for any operator T and every set Ω for which von Neumann's inequality holds?
That is certainly the case for self-adjoint matrices as can be derived by Theorem 12, Theorem 4 and the definition of the spectral norm. In fact, self-adjoint matrices have a so-called, spectral decomposition. This entails that a self-adjoint matrix T may be rewritten as the sum of each of its eigenvalues times the orthogonal projection onto the eigenvalue space. In other words, if T is a self-adjoint matrix with eigenvalues $\lambda_1, \lambda_2, ...$, then

$$T = P_{\lambda_1}\lambda_1 + P_{\lambda_2}\lambda_2 + ...,$$

which is referred to as the spectral decomposition of T. By this spectral decomposition we may see that the norm of a Hermitian operator is equal to the spectral radius of T, $\sigma_{max}(T)$. As we have seen from Theorem 12: let T be any operator acting on an infinite dimensional Hilbert space and let p be any polynomial with complex coefficients, then:

$$\|p(T)\| \leq \sup_{|z| \leq \|T\|} |p(z)|. \tag{9}$$

The norm of z must hence be smaller or equal to that of T, but the norm of T is equal to $\sigma_{max}(T)$, the modulus of the largest eigenvalue of T. Since $|z| \leq \|T\|$ describes a disk with radius $\|T\|$, this disk comprises all eigenvalues of T and Ω must at least be equal to this disk, hence $\Omega \supseteq \sigma(T)$.
Before continuing, we must make note of the fact that the set $\sigma(T)$ is defined in a slightly different fashion for operators than for matrices. In fact, the spectrum of an operator contains more than just the eigenvalues of that operator. This does not impact what has been said before, but we will include, for clarity, this definition as well.

Definition 6.1 (Spectrum for operators)**.** *Let T be a linear operator acting on a Hilbert space H over the field $\mathbb{C}$ and let I be the identity operator on H. The spectrum of T, $\sigma(T)$, is the set of all $\lambda \in \mathbb{C}$ for which the operator $\lambda I - T$ does not have an inverse that is a bounded linear operator.*

As we have just seen, the set Ω for self-adjoint operators must clearly contain the spectrum of the operator taken into consideration and, in fact, that is the case for all operators as follows from the following theorem.

Theorem 15. *Let T be any operator acting on a Hilbert space H and let Ω be a closed bounded set such that for any polynomial p*

$$\|p(T)\| \leq c \sup_{z \in \Omega} |p(z)|,$$

where $c \in \mathbb{R}$ is a scalar independent of p. Then $\Omega \supseteq \sigma(T)$.

Proof. Suppose there exists some element $\lambda \in \sigma(T)$ that is not contained in Ω. Since $\lambda \notin \Omega$, the following inequality must hold for $f(z) = \frac{1}{\lambda - z}$, seeing as f may be written as a uniform limit of polynomials on Ω:

$$\|(\lambda I - T)^{-1}\| \leq c \sup_{z \in \Omega} |\frac{1}{\lambda - z}|.$$

This implies that the inverse $(\lambda I - T)^{-1}$ of the operator $\lambda I - T$ exists as a bounded operator. This clearly contradicts Definition 6.1, since we had assumed λ to be an element of $\sigma(T)$. It must then be the case that $\lambda \in \sigma(T)$ implies that $\lambda \in \Omega$. □

Now that we have established that $\Omega \supseteq \sigma(T)$, we must ask ourselves whether or not $\Omega = \sigma(T)$ is a suitable candidate. We will show that is not the case through a counterexample.
Let T be an operator of the following form:

$$T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This operator T has eigenvalue 0 with multiplicity of 2. This implies that

$$\sigma_{max}(T) = \max_{1 \leq i \leq 2} |\lambda_i| = \max\{|0|\} = |0| = 0.$$

Furthermore, $\sigma(T) = \{0\}$. Suppose now that indeed $\Omega = \sigma(T)$. Then the following inequality should hold for any polynomial p:

$$\|p(T)\| \leq |p(0)|.$$

This does not hold for just any polynomial p. Take for example $p = z$. This yields the following inequality:

$$\left\| \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\| = 1 \leq 0 = c \sup_{z \in \{0\}} z$$

Clearly the inequality does not hold, not even for large values of c. Hence it can not be the case that $\Omega = \sigma(T)$.

6.2 The numerical range

As it turns out, a good candidate for Ω is the numerical range of an operator.

Definition 6.2 (Numerical range of an operator)**.** *Let T be an operator acting on a complex Hilbert space, then the numerical range of T, denoted as $W(T)$, is defined as follows:*

$$W(T) = \{(Tx, x) \in \mathbb{C} : \|x\| = 1\}.$$

This is clearly a smaller set than $\overline{\mathbb{D}}$ if T is a contraction, as we will later see from Lemma 17, but does it then also provide a lower upper bound for von Neumann's inequality?
Before answering this question, it is of importance we state and prove certain properties belonging to the numerical range of our operator T. First of all we shall show that $W(T)$ is a convex set following a proof by Karl Gustafson of the Toeplitz-Hausdorff Theorem, [7].

Theorem 16 (Toeplitz-Hausdorff)**.** *The numerical range $W(T)$ of an arbitrary linear operator T in a Hilbert space is convex.*

Proof. It suffices to consider the situation $(Tx_1, x_1) = 0$, $(Ax_2, xz_2) = 1$, $\|x_i\| = 1$ $(i = 1, 2)$. This because $W(\mu T + \gamma I) = \mu W(T) + \gamma$ for scalars $\mu, \gamma \in \mathbb{C}$.
Let $x = \alpha x_1 + \beta x_2$, $\alpha, \beta \in \mathbb{R}$, and require that

$$\|x\|^2 = \|\alpha x_1 + \beta x_2\|^2 = \alpha^2 + \beta^2 + 2\alpha\beta Re(x_1, x_2) = 1 \tag{10}$$

and that for each λ, $0 \leq \lambda \leq 1$

$$(Ax, x) = \beta^2 + \alpha\beta\{(Ax_1, x_2) + (Ax_2, x_1)\} = \lambda \tag{11}$$

If $(Ax_1, x_2) + (Ax_2, x_1) \in \mathbb{R}$, then the system described by Equations (10) and (11) clearly possesses solutions.
In fact, $(Ax_1, x_2) + (Ax_2, x_1)$ can always be guaranteed to be real by using an appropriate (scalar multiple of) x_1. □

Secondly, we shall show that the numerical range of an operator T is a subset of the ball of radius $\|T\|$ about the origin.

Lemma 17. *The numerical range of an operator T, $W(T)$, is a subset of the disk $|z| \leq \|T\|$.*

Proof. Let $z = (Tx, x)$ for $\|x\| = 1$. Then

$$|z| = |(Tx, x)| = |(x, Tx)| \leq \|x\|\|Tx\| \leq \|x\|^2\|T\| = \|T\|.$$

Clearly then $|z| \leq \|T\|$ for all $z \in W(T)$. This proves that $W(T)$ is a subset of the disk of radius $\|T\|$ centred at 0. □

It is this second property in particular which eventually led to the result proved by Okubo and Ando, [1]. Okubo and Ando proved that $\mathbb{D}$ is a complete c-spectral set for a contraction T. In other words, they proved that $\mathbb{D}$ is a set Ω for which the equality

$$\|p(T)\| \leq c \sup_{z\in\Omega} |p(z)|$$

holds.

Theorem 18 (Okubo-Ando Inequality)**.** *Let T be a contraction acting on a Hilbert space H. If $W(T) \subseteq \overline{\mathbb{D}}$, then the following equation holds for all polynomials p with complex coefficients:*

$$\|p(T)\| \leq 2 \sup_{z \in \mathbb{D}} |p(z)|.$$

One may already notice that, while this inequality holds at the cost of $c = 2$, it also relies on the weaker assumption that $W(T) \subseteq \overline{\mathbb{D}}$.

6.3 Crouzeix's Conjecture

Theorem 18 paved the path to a more general inequality which could hold for all operators T and their numerical range $W(T)$. Such an inequality was produced by Michel Crouzeix, who in 2006 produced a paper in which he proved this more general inequality would hold for $c = 11.08$, [5].

Theorem 19 (Crouzeix's Theorem)**.** *Let T be an operator acting on a Hilbert space H and let p be any polynomial with complex coefficients. Then*

$$\|p(T)\| \leq 11.08 \sup_{z \in W(T)} |p(z)|.$$

Once again, while this inequality holds at the great cost of $c = 11.08$, it also relies on the weaker assumption, making the statement weaker as a whole. This bound has though since changed as in 2017 Crouzeix and Palencia found and proved that there exists a much better upper bound for this inequality, [4].

Theorem 20. *Let T be an operator acting on a Hilbert space H and let p be any polynomial with complex coefficients. Then*

$$\|p(T)\| \leq (1 + \sqrt{2}) \sup_{z \in W(T)} |p(z)|.$$

Thus far, this has been the lowest bound to be found and proven, although Michel Crouzeix also released a paper in 2004 in which he supposed there to an even lower bound, [3].

Conjecture 1 (Crouzeix's conjecture)**.** *Let $T \in \mathbb{C}^{n \times n}$ be an operator acting on a Hilbert space H and let p be any polynomial with complex coefficients. Then*

$$\|p(T)\| \leq 2 \sup_{z \in W(T)} |p(z)|.$$

Unfortunately, Crouzeix's conjecture has thus far only been proven for $n = 2$ and still an open problem for $n > 2$. One thing is certain, for $\Omega = W(T)$ there is no lower bound possible than $c = 2$. This can be easily seen if we consider the operator

$$T = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$$

and $p(z) = z$. Then

$$\|p(T)\| = \left\| \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \right\| = 2 \leq c \sup_{z \in W(T)} |p(z)| = c \sup_{z \in W(T)} |z| = c.$$

For this inequality to hold, clearly c must at least be equal to 2, hence no lower bound can be found. Once again, this could inequality holds at cost of $c = 2$, but once more relies on the weaker assumption, making the statement weaker as a whole.

7 Conclusion

As we have seen throughout this thesis, although von Neumann's inequality's proof has origin in a mathematical field very different to the one we used to approach it, it still stands when backed up by Linear Algebra. In particular, we should mention that it is rather surprising that the Maximum Modulus Principle, Theorem 7, which belongs to and finds its use in Complex Analysis, may also be proven by means of Linear Algebra. The possibility to achieve this is really what makes it possible to let von Neumann's Inequality rest solely on Linear Algebra, while preserving all of its properties. In fact, the result attained in this thesis was only possible thanks to the combination of certain already existing techniques attributed to Orr Moshe Shalit, [11], and Gilles Pisier, [9]. In order to do this, we first recalled all needed notions and prerequisites, such as the definition of a contraction or the Spectral Theorem. We then proved the Maximum Modulus Principle by means of Linear Algebra and used this result in proving von Neumann's Inequality. In addition, we also peeked into the classical proof of von Neumann's Inequality and into variations of this inequality. Of particular interest is that Crouzeix's conjecture, Conjecture 1, is still an open problem on which more research could be done. Furthermore, Okubo-Ando's Inequality, Theorem 18, offers options for further research in line with what has been achieved in this thesis: would it be possible to prove Okubo-Ando's Inequality only utilising techniques belonging to Linear Algebra?

As a final remark, we would like to briefly touch upon the applications of inequalities similar to von Neumann's. Such inequalities may be used to bound the error when solving huge Linear Systems, [8].

References

[1] Catalin Badea, Michel Crouzeix, and Hubert Klaja. “Spectral sets and operator radii”. In: *Bulletin of the London Mathematical Society* 50 (2018), pp. 986–996. DOI: 10.1112/blms.12191.

[2] Garrett Buffington. “Polar Decomposition of a Matrix”. In: (2014). URL: http://buzzard.ups.edu/courses/2014spring/420projects/math420-UPS-spring-2014-buffington-polar-decomposition.pdf.

[3] M. Crouzeix. “Bounds for Analytical Functions of Matrices”. In: *Integral Equations and Operator Theory* 48 (2004), pp. 461–477. DOI: 10.1007/s00020-002-1188-6.

[4] M. Crouzeix and C. Palencia. “The Numerical Range is a $(1+\sqrt{2})$-Spectral Set”. In: *Journal on Matrix Analysis and Applications* 38.2 (2017), pp. 649–655. DOI: 10.1137/17M1116672.

[5] Michel Crouzeix. “Numerical range and functional calculus in Hilbert space”. In: *Journal of Functional Analysis* 244 (2006), pp. 668–690. DOI: 10.1016/j.jfa.2006.10.013.

[6] Bernard Friedman. *Principles and techniques of applied mathematics*. Courier Dover Publications, 1990.

[7] Karl Gustafson. “The Toeplitz-Hausdorff Theorem for linear operators”. In: *American Mathematical Society* (1969), pp. 203–204. URL: https://www.ams.org/journals/proc/1970-025-01/S0002-9939-1970-0262849-9/S0002-9939-1970-0262849-9.pdf.

[8] Martin H. Gutknecht. “A Brief Introduction to Krylov Space Methods for Solving Linear Systems”. In: (2007). DOI: 10.1007/978-3-540-46375-7_5.

[9] Gilles Pisier. *Similarity Problems and Completely Bounded Maps*. New York: Springer-Verlag Berlin Heidelberg, 1996. Chap. 1. DOI: 10.1007/978-3-662-21537-1.

[10] Edward B. Saff and Arthur David Snider. *Fundamentals of Complex Analysis Engineering, Science, and Mathematics*. Essex: Pearson Education Limited, 2014. Chap. 6.

[11] Orr Moshe Shalit. “A sneaky proof of the maximum modulus principle”. In: *The American Mathematical Monthly* 120.4 (2013), pp. 359–362. DOI: 10.4169/amer.math.monthly.120.04.359.