# The Krein-Milman theorem and its applications

*A Thesis Submitted in Fulfillment of the Requirements for the Degree of*
Bachelor in Applied Mathematics

*by*

**Siem van Benthem**

Supervised by: Mario Klisse and Martijn Caspers

Faculty of Electrical Engineering, Mathematics and Computer Science
DELFT UNIVERSITY OF TECHNOLOGY

June 2021

# Contents

# 1 Introduction

The first ever study on extreme points dates back to Minkowski in 1911. Minkowski [1] examined the relation connecting convex sets in three-dimensional spaces, and their set of extreme points. His research uncovered that a compact and convex set (in a three-dimensional space) coincides with the convex hull of the set of its extreme points. His work inspired other mathematicians to explore whether extreme points of convex sets are associated likewise in a more general setting. The most significant research regarding this speculation is due to the mathematicians Mark Krein and David Milman in the year 1940. The result presented in their paper [2] is known as the Krein-Milman theorem, and is the most recognized statement regarding convex sets within functional analysis. It extends Minkowski's result from three dimensional spaces, to locally convex Hausdorff spaces.

As with any theorem or statement in mathematics, it is natural to speculate in what manner the Krein-Milman theorem may be adopted to our advantage. After the publishment of the paper by Krein and Milman, mathematicians have employed the Krein-Milman theorem in various forms. De Branges [3] applies the Krein-Milman theorem to prove the notorious Stone-Weierstrass theorem, which states (in short) that a separating subspace of the continuous functions is dense. In the majority, the Krein-Milman theorem is utilized to reduce the problem of identifying what elements belong to some set, to finding the extreme points of this set. Kadets [4] and Phelps [5] utilize this technique to prove Berstein's theorem, which describes completely monotone functions as Laplace transforms of non-negative Borel measures. Mauldon [6] proves Kendall's theorem, which states that the extreme points of the space of (possibly infinite) doubly stochastic matrices are exactly the permutation matrices.

The objective of this thesis is to study the various applications of the Krein-Milman theorem, within numerous fields of mathematics. The first part of the thesis familiarizes the reader with several necessary definitions, theorems and other terminology. Subsequently, the statement of the Krein-Milman and its proof are introduced. Then, numerous applications of the Krein-Milman theorem in functional analysis, integration theory, linear algebra and graph theory are studied. The thesis is concluded with a discussion on the assumptions of the Krein-Milman theorem, where compactness is largely considered. Klee [7] and Nachbin [8] consider families of (not necessarily compact) convex sets satisfying a statement similar to the Krein-Milman theorem. Roberts [9] provides a complex counterexample to the Krein-Milman theorem if the space is not locally convex.

# 2 Preliminaries and basics

This section covers the basic notions regarding the theory of topological vector spaces that will be used in the thesis. Throughout this section let $E$ be a topological vector space over the field of real or complex numbers. Many definitions introduced in this section will be familiar to the reader, however we repeat them for the purpose of completeness.

## 2.1 Convex sets and extreme points

In defining the convexity of sets, the notion of a line segment or interval is required. For any two points $a, b \in E$, the *open line segment* between $a$ and $b$ is the collection of all convex combinations $\lambda a + (1-\lambda)b$ for $\lambda \in (0, 1)$. Note that the open line segment excludes the so-called *endpoints* $a$ and $b$. Similarly, the *closed line segment* connecting $a$ and $b$ is the union of the open line segment and its endpoints. Notationwise, it is customary to write $(a, b)$ for the open interval between $a$ and $b$, and $[a, b]$ for the closed one. Now one can define convex sets.

**Definition 2.1.** *A subset $A \subset E$ is called convex if $[a, b] \subset A$ for all $a, b \in A$.*

Intuitively, one can think of convex sets as ones that contain no holes or inlets. Such sets have many interesting properties, some of which are the Krein-Milman and the Hahn-Banach theorem. An important operation for the Krein-Milman theorem is taking the convex hull of a given set. For a set $A \subset E$, the *convex hull* of $A$, denoted by conv $A$ is the intersection of all convex subsets of $E$ containing $A$. Thus conv $A$ is the smallest convex set containing $A$. Observe, that any intersection of convex sets is convex again. In (possibly non-convex) sets, there exist points that are not contained in any line segment connecting two other points. It turns out that knowledge about these points gives us information of the entire set, whenever the set in question is convex. These points are the so-called extreme points.

**Definition 2.2.** *Let $A$ be a convex and nonempty subset of $E$.*

(i) *An element $x \in A$ is called an extreme point of $A$ if $x \notin (a, b)$ for all $a, b \in A$.*

(ii) *A set $X \subset A$ is called a face or an extreme subset of $A$ whenever the following implication is valid: if $\lambda a + (1-\lambda)b \in X$ for some $a, b \in A$ and $\lambda \in (0, 1)$, then $a, b \in X$.*

The set of all extreme points of some set $A$ will be denoted by ext $A$. Extreme points can also be defined for non-convex sets. In this case, a point $x \in A$ is called an extreme point if it is not contained in any open line segment contained in $A$. A consequence of the definition of a face is the following

**Proposition 2.3.** *Let $A \subset E$ be a convex set and $\mathcal{X}$ any collection of faces of $A$. Then $\bigcap \mathcal{X}$ is also a face of $A$.*

*Proof.* Observe that the empty set is always a face of a convex set, thus assume $\bigcap \mathcal{X} \neq \varnothing$. Let us argue by contradiction, thus assume that $\bigcap \mathcal{X}$ is not a face of $A$. Then there exists an element $x \in \bigcap \mathcal{X}$ and $a, b \in A$ where $a \notin \bigcap \mathcal{X}$ such that $x \in (a, b)$. As a consequence there must exist a face $X \in \mathcal{X}$ that does not contain $a$, otherwise $a \in \bigcap \mathcal{X}$. Regardless of where the point $b$ lies, it follows that the set $X$ is not a face since $x \in X$ is contained in a line segment of which at least one endpoint lies outside of $X$. This is a contradiction since $X$ was assumed to be a face, which implies that $\bigcap \mathcal{X}$ is a face of $A$. $\square$

## 2.2 Locally convex spaces, semi-norms and the Hausdorff property

Recall the definition of the base of a topology. A *base* of a topology $\tau$ on $E$ is a collection of open sets $\mathcal{B} \subset \tau$ such that any set in $\tau$ is the union of open sets in $\mathcal{B}$. A locally convex topological vector space is one with a specific kind of base.

**Definition 2.4.** *A topological vector space is called locally convex if there exists a base of convex neighborhoods of zero.*

From here onwards, a locally convex topological vector space will be referred to as a locally convex space. Such spaces can also be defined differently. Recall first some terminology on semi-norm induced topologies. For a semi-norm $p$ on $E$, define the open ball of $p$ with radius $\varepsilon$ as $B_{p,\varepsilon} = \{x \in E \,:\, p(x) < \varepsilon\}$. For a family $\mathcal{P}$ of semi-norms on $E$, let $\mathcal{B}_{\mathcal{P}}$ be the collection of all finite intersections of open balls $\mathcal{B}_{p,\varepsilon}$, with $p \in \mathcal{P}$ and $\varepsilon > 0$. The *topology induced by $\mathcal{P}$ on $E$* is the topology with base $\mathcal{B}_{\mathcal{P}}$.

It turns out that the topology on a locally convex space over the complex or real numbers is always induced by a family of semi-norms [10]. These semi-norms can be used to characterize convenient properties of topological vector spaces. Several of these will be presented here. Let us start with the Hausdorff property. Recall that a topological vector space is called *Hausdorff* if for any two distinct elements $a, b \in E$ there exist disjoint open neighborhoods $A$ of $a$, and $B$ of $b$. The next proposition demonstrates the convenient criterion for the Hausdorff property in locally convex spaces.

**Proposition 2.5.** *Let $E$ be locally convex and $\mathcal{P}$ the family of semi-norms inducing the topology on $E$. Then $E$ is Hausdorff if and only if for all $a \in E$, there is a semi-norm $p \in \mathcal{P}$ such that $p(a) \neq 0$.*

For a proof of this proposition, we would like to refer to [10]. The condition on $\mathcal{P}$ in the proposition is often referred to as the separation condition. If a family of semi-norms satisfies this condition, then it is called *separated*. Proceeding, let us consider a characterization of continuity of linear maps between locally convex spaces. A proof of this criterion can be found in [11].

**Proposition 2.6.** *Let $E$ and $F$ be locally convex spaces, whose topologies are induced by the totally ordered families of semi-norms $\mathcal{P}$ and $\mathcal{Q}$ respectively. The linear map $L : E \to F$ is continuous if and only if for every $q \in \mathcal{Q}$, there exists a $p \in \mathcal{P}$ and $M > 0$ such that for all $x \in E$*

$$q(\,L(x)\,) \leq M\,p(x)$$

In locally convex spaces, boundedness of sets is completely prescribed by the values attained by the semi-norms on this set, which is demonstrated by the next proposition.

**Proposition 2.7.** *Let $E$ be a locally convex space whose topology is induced by the family of semi-norms $\mathcal{P}$. A subset $A$ of $E$ is bounded if and only if $p(A)$ is bounded for all $p \in \mathcal{P}$.*

A proof can be found in [11]. Conveniently, with the aid of this proposition, showing boundedness of sets reduces to deriving finiteness of the semi-norms on this set. Lastly, there exists a beneficial convergence characterization with respect to the semi-norms.

**Proposition 2.8.** *Let $E$ be a locally convex space whose topology is induced by the family of semi-norms $\mathcal{P}$, that contains the net $(f_\tau)_{\tau \in T}$. Then $f_\tau \to f$ if and only if $p(f_\tau - f) \to 0$ for all $p \in \mathcal{P}$.*

The reader can consult [11] for a proof of this statement.

## 2.3 Weak*-topology, product topology and the Banach-Alaoglu theorem

Two topologies will be introduced in this section. These are the product and the weak*-topology. The main benefit of these topologies is that they are well-behaved with respect to compactness of their subsets. Both of these topologies provide a sufficient structure for checking compactness quite effortlessly. For sets equipped with the product topology, the Tikhonov theorem will be our main tool, and for the weak*-topology this will be the Banach-Alaoglu theorem.

First we introduce some necessary terminology. Let $E_t$ be a family of real or complex topological vector spaces, where $t \in T$ is an index set. Further, define $E$ to be the infinite direct product of the sets $E_t$. On the set $E$ define the *canonical projections* $\pi_t : E \to E_t$ which map an element in $E$ to its $t$-th coordinate. With this terminology the product topology can be defined.

**Definition 2.9** (Product topology)**.** *Let $E_t$ $(t \in T)$ be a family of topological spaces and let $E$ be the direct product of the sets $E_t$. The product topology on $E$ is the coarsest topology for which all canonical projections $\pi_t$ are continuous.*

Equipping a set with the product topology turns it into a topological vector space. A set equipped with the product topology is often called a *product space*. A product space is Hausdorff if each of the topological spaces in the product are Hausdorff. The same holds for local convexity, if each of the spaces in the product are locally convex, then the product space is locally convex as well. The product topology often goes by another name, the *Tikhonov-topology*. A result related to the product topology is Tikhonov's theorem. It states that compactness of a set in each of the topological spaces $E_t$ of a product space $E$, is equivalent to compactness in $E$ itself with respect to the product topology. The formal statement is given below.

**Theorem 2.10** (Tikhonov theorem)**.** *The product space $E = \prod_{t \in T} E_t$ is compact if and only if $E_t$ is compact for all $t \in T$.*

A proof of the Tikhonov theorem may be found in [4]. Let us shift our attention from the product topology to the weak*-topology, which will be defined in this section as well. Before introducing this definition, the concept of a dual space is necessary.

**Definition 2.11** (Dual space)**.** *The dual space $E^*$ of $E$ is the set of all continuous linear functions $f : E \to \mathbb{K}$.*

A function that belongs to the dual space is called a *functional*. A conventional norm on the dual space is the *operator norm*. The operator norm can be defined for general spaces of continuous linear maps between normed vector spaces, but we chose to define it for the dual space of a normed vector space only.

**Definition 2.12** (Operator norm)**.** *Let $E$ be a normed space with norm $||\cdot||$. The operator norm $||\cdot||_{\text{op}}$ on the dual space $E^*$ is the norm defined by*

$$||\, f \,||_{\text{op}} = \sup\{\, |\, f(x)\,| \,:\, x \in E,\ ||\, x\,|| \leq 1\,\}.$$

So the operator norm evaluates the supremum of a function attained in the unit ball in $E$. Conveniently, equipping the dual space with the operator norm and the weak*-topology (which will be defined shortly after), provides a structure that allows the unit ball to be compact, as demonstrated by the Banach-Alaoglu theorem. The convenience of the weak*-topology is encouraged even more when considering the fact that the unit ball is not compact in any infinite dimensional normed space, with respect to the norm-induced topology. There are many (equivalent) ways to define the weak*-topology, but in this thesis we use the definition regarding the continuity of evaluation functionals.

**Definition 2.13** (Weak*-topology)**.** *The weak*-topology on the dual space $E^*$ is the coarsest topology such that the evaluation functionals $\delta_x : E^* \to \mathbb{K}$ defined by $\delta_x(f) = f(x)$, are continuous for all $x \in E$.*

For an explicit construction of the open sets in the weak*-topology topology, we would like to refer to [10]. Equipping a dual space with the weak*-topology, turns it into a locally convex Hausdorff space. A convenient characterization of convergence in the weak*-topology emerges from the definition, which is demonstrated by the next proposition.

**Proposition 2.14.** *Let $(f_\tau)_{\tau \in T}$ be a net in $E^*$. Then $f_\tau \to f$ with respect to the weak*-topology if and only if $f_\tau(x) \to f(x)$ for all $x \in E$.*

*Proof.* Suppose first that $f_\tau \to f$ with respect to the weak*-topology. Then for all $x \in X$ the evaluation functionals $\delta_x$ are continuous and therefore $\delta_x(f_\tau) \to \delta_x(f)$ in $\mathbb{K}$, which is exactly pointwise convergence. It is easily checked that the inverse mappings of the evaluation functionals are continuous also, since we are free to choose a topology on $\mathbb{K}$ (for example the discrete topology). Therefore, by a similar argument, if $f_\tau(x) \to f(x)$ for all $x \in E$, then $f_\tau \to f$ with respect to the weak*-topology. $\square$

Thus a net convergences with respect to the weak*-topology if and only if it converges pointwise in $\mathbb{K}$. Now we are ready for the Banach-Alaoglu theorem, which is stated without proof. The interested reader can find a proof in [4], though most books on functional analysis supply a proof of this statement.

**Theorem 2.15** (Banach-Alaoglu)**.** *Let $E$ be a normed space and equip $E^*$ with the operator norm. Then the closed unit ball in $E^*$ is compact with respect to the weak*-topology.*

## 2.4 The Riesz-Markov-Kakutani representation theorem and measure theory

The main result introduced in this section is the Riesz-Markov-Kakutani representation theorem, or shorter the Riesz-representation theorem. Additionally, some necessary terminology from measure theory will be introduced. Throughout this section, let $(S, \Sigma)$ be a measurable space. Let us first deal with Borel measures.

Let $(E, \tau)$ be a topological space and $\Sigma$ some $\sigma$-algebra on $E$. A measure $\mu$ on the measurable space $(E, \Sigma)$ is called *inner regular* if

$$\mu(A) = \sup\{\, \mu(B) \,:\, B \subset A, \text{ compact and measurable}\,\},\ A \in \Sigma.$$

Similarly, the measure $\mu$ is called *outer regular* whenever

$$\mu(A) = \inf\{\, \mu(B) \,:\, B \supset A, \text{ open and measurable}\,\},\ A \in \Sigma.$$

A measure that is both inner regular and outer regular is called a *regular measure*. Proceeding, let $\mathcal{B}(\tau)$ be the smallest $\sigma$-algebra containing all of the sets in $\tau$. This $\sigma$-algebra will be referred to as the *Borel $\sigma$-algebra with respect to $\tau$*. A measure on a topological space $E$ is called a *Borel measure* when it is a measure on the measurable space $(E,\, \mathcal{B}(\tau)\,)$. Furthermore, such a measure $\mu$ is called a *probability measure* if it is also regular and satisfies $\mu(E) = 1$.

In addition to measures that take values in $[0, \infty]$, we will also consider *signed measures* and *complex measures*. A signed measure or *charge* defined on the measurable space $(S, \Sigma)$ is a function $\mu : \Sigma \to \mathbb{R}$ that is $\sigma$-additive and satisfies $\mu(\varnothing) = 0$. Signed measures admit a practical decomposition, which allows us to write them as a linear combination of non-negative measures. This is called the *Hahn-Jordan decomposition*. Before introducing this result, we consider a closely related result, called the Hahn-decomposition theorem. A proof of this statement can be found in [4].

**Theorem 2.16** (Hahn-decomposition)**.** *Let $\mu$ be a signed measure on the measurable space $(S, \Sigma)$. Then there exist measurable sets $S_1$ and $S_2$ such that*

(i) $S_1 \cup S_2 = S$ *and* $S_1 \cap S_2 = \varnothing$*;*

(ii) $\mu(B) \geq 0$ *for all measurable* $B \subset S_1$*;*

(iii) $\mu(B) \leq 0$ *for all measurable* $B \subset S_2$*.*

*Moreover, this decomposition is unique op to null sets.*

It is customary to write $(S_1, S_2)$ for the Hahn-decomposition of a measure space. Given a signed measure defined on a measurable space $(S, \Sigma)$, the Hahn-decomposition can be adopted to create a decomposition of the signed measure $\mu$. This decomposition is given in the next statement and is, as mentioned priorly, referred to as the Jordan or the Hahn-Jordan decomposition.

**Theorem 2.17** (Hahn-Jordan decomposition)**.** *Let $\mu$ be a signed measure on the measurable space $(S, \Sigma)$ with Hahn-decomposition $(S_1, S_2)$. Then $\mu$ has a unique decomposition $\mu = \mu^+ - \mu^-$ such that*

(i) $\mu^+$ *and* $\mu^-$ *are non-negative;*

(ii) $\mu^+(B) = 0$ *for all measurable* $B \subset S_2$*;*

(iii) $\mu^-(B) = 0$ *for all measurable* $B \subset S_1$.

A proof of the Hahn-Jordan decomposition theorem can be found in the same reference of the Hahn-decomposition theorem. Proceeding, for a signed measure $\mu$, we define the *variation measure* using the Hahn-Jordan decomposition as

$$|\mu| = \mu^+ + \mu^-.$$

When dealing with signed measures, rather than the measure itself, the variation measure is often considered instead. Additionally, define the *total variation* $||\,\mu\,||$ of a signed measure $\mu$ on $(S, \Sigma)$ as the value of $|\mu|(S)$. When integrating a measurable function $f$ with respect to the variation measure, it is customary to write $\int f\, |d\mu|$.

A function $\mu : \Sigma \to \mathbb{C}$ on $(S, \Sigma)$ is called a *complex measure* if it is $\sigma$-additive and satisfies $\mu(\varnothing) = 0$. Similar to complex valued functions, complex measures can be decomposed into a real and an imaginary part. Intuitively, for a complex measure $\mu$ and measurable set $B \in \Sigma$, define the real part and complex parts as $(\operatorname{Re}\mu)(B) = \operatorname{Re}(\,\mu(B)\,)$ and $(\operatorname{Im}\mu)(B) = \operatorname{Im}(\,\mu(B)\,)$ respectively. Consequently, it is obvious that for all $B \in \Sigma$

$$\mu = \operatorname{Re}\mu + i \operatorname{Im}\mu.$$

Observe that the real and imaginary part are signed measures, which can be decomposed into non-negative measures due to the Hahn-Jordan decomposition. Hence complex measures can be decomposed into non-negative measures. For complex measures, the variation measure can be defined as well. However, this is different from the definition of signed measures. Let $B \in \Sigma$ and let $\pi(B)$ denote the family of all finite pairwise disjoint collections of measurable subsets of $B$. Then the variation measure of the complex measure $\mu$ is given by

$$|\mu|(B) = \sup_{\mathcal{B} \in \pi(B)} \sum_{A \in \mathcal{B}} |\,\mu(A)\,|.$$

One can check that both definitions for the total variation measure coincide whenever $\mu$ is a signed measure. Again, when dealing with complex measures, rather than the measure itself, the variation measure is often considered instead.

At this moment we will shift our focus towards the Riesz-Markov-Kakutani representation theorem. Similar to the Hahn-Banach theorems, there are many statements similar to the one given here. The Riesz-Markov-Kakutani representation theorem associates continuous linear functionals with complex Borel measures.

Recall that a function $f \in C(E, \mathbb{C})$ is said to *vanish at infinity* if for all $\varepsilon > 0$ there exists a compact set $A \subset E$ such that $|\, f(x)\,| < \varepsilon$ whenever $x \notin A$. Denote by $C_0(E, \mathbb{C})$ the space of functions $f \in C(E, \mathbb{C})$ which vanish at infinity. Lastly, the space $E$ is called *locally compact* if every element $x \in E$ has a compact neighborhood.

**Theorem 2.18** (Riesz-Markov-Kakutani)**.** *If $E$ is locally compact and Hausdorff and $\psi \in C_0(E, \mathbb{C})^*$, then there exists a unique regular complex Borel measure $\mu$ such that*

$$\psi(f) = \int_E f \, d\mu, \quad \text{for all } f \in C_0(E, \mathbb{C}).$$

In addition, if the functional $\psi$ in the statement is positive, then $\mu$ is a non-negative measure. For a proof of the Riesz-Markov-Kakutani representation theorem we refer to [12]. There are many convenient consequences of the Riesz-representation theorem, as we will now demonstrate.

For a topological vector space $E$, denote by $\mathcal{B}$ the vector space of all regular complex Borel measures on $E$. Here we consider $C_0(E, \mathbb{C})^*$ with respect to the weak*-topology. As a consequence of the Riesz-representation theorem, there exists a bijection $\rho : \mathcal{B} \to C_0(E, \mathbb{C})^*$, that is defined by $(\,\rho(\mu)\,)(f) = \int_E f \, d\mu$. Indeed, this mapping is injective due to the uniqueness in the theorem. Additionally, surjectivity is a direct consequence of the Riesz-representation theorem.

Additionally, using the Riesz-representation theorem we can define the weak*-topology on the space of measure $\mathcal{B}$. The weak*-topology on $\mathcal{B}$ is the unique topology on $\mathcal{B}$ such that $\rho$ is a homeomorphism. The open sets of this topology are exactly the images of the open sets of the weak*-topology on $C_0(E, \mathbb{C})^*$, under the bijection $\rho$. Recall the convenient convergence criterion in $C_0(E, \mathbb{C})^*$ (with respect to weak*-topology) from Proposition 2.14. A similar result holds for the weak*-topology on $\mathcal{B}$, namely that a net $(\mu_\tau)_{\tau \in T}$ converges to $\mu$ with respect to the weak*-topology if and only if $\int_E f \, d\mu_\tau$ converges to $\int_E f \, d\mu$ for all $f \in C_0(E, \mathbb{C})$.

## 2.5 Other required results

This final section on the preliminaries and basics introduces the last required results that are necessary for the thesis. This includes one of the Hahn-Banach theorems. First we recall a proposition from set theory, namely the Kuratowski-Zorn lemma.

A set $P$ that is equipped with a partial ordering $\preceq$ is called a *partially ordered* set. For this, one usually writes $(P, \preceq)$. For the corresponding definitions on partial orderings we refer to [10]. A subset $C$ of a partially ordered set $P$ is called a *chain* if for any two elements $a, b \in C$ either $a \preceq b$, or $b \preceq a$. Such sets are also referred to as linearly ordered or totally ordered sets. Lastly, a *majorant* or *upper bound* of a chain $C$ is an element $a \in P$ such that $b \preceq a$ for all $b \in C$. The Kuratowski-Zorn lemma admits a criterion for the existence of maximal elements in the set $P$ with respect to the partial ordering $\preceq$. A proof of this statement can be found in nearly all books on set theory.

**Lemma 2.19** (Kuratowski-Zorn)**.** *If every chain in a partially ordered set $(P, \preceq)$ has a majorant, then $P$ has maximal elements.*

Continuing, we shift our attention from set theory back to functional analysis, and in particular to the Hahn-Banach theorem, which forms a fundamental result in the theory. The proof of the Hahn-Banach relies strongly on the Kuratowski-Zorn lemma. The Hahn-Banach theorem stated here is also known as the Hahn-Banach separation theorem or the geometric Hahn-Banach theorem.

**Theorem 2.20** (Hahn-Banach)**.** *Let $A, B \subset E$ be nonempty convex and disjoint.*

- *If $A$ is open, then there exists a continuous linear functional $f$ such that* $\operatorname{Re} f(a) < \operatorname{Re} f(b)$ *for all $a \in A$ and $b \in B$. Furthermore,* $\sup_{x \in A} \operatorname{Re} f(x) \leq \inf_{x \in B} \operatorname{Re} f(x)$.
- *If $E$ is locally convex, $A$ is compact and $B$ closed, then there exists a continuous linear functional $f$ such that* $\operatorname{Re} f(a) < \operatorname{Re} f(b)$ *for all $a \in A$ and $b \in B$. Furthermore,* $\sup_{x \in A} \operatorname{Re} f(x) < \inf_{x \in B} \operatorname{Re} f(x)$.

A proof of the Hahn-Banach theorem presented here can be found in [11]. Another well-known result is the extreme value theorem, which we do not prove.

**Theorem 2.21** (Extreme value theorem)**.** *If $\mathbb{K} = \mathbb{R}$, $A \subset E$ is compact and $f : E \to \mathbb{K}$ continuous, then $f$ is bounded and attains its supremum and infimum on $A$.*

Proceeding, we state and prove two results regarding compact sets.

**Lemma 2.22.** *Any compact subset of a Hausdorff space is also closed.*

*Proof.* Let $A \subset E$ be compact and choose $x \in A^c$. Observe that, as $E$ is Hausdorff, for any $y \in A$ there exist open neighborhoods $N_{(x,y)}$ of $x$ and $M_{(x,y)}$ of $y$, such that $N_{(x,y)} \cap M_{(x,y)} = \varnothing$. The family $\{M_{(x,y)} | y \in A\}$ of open neighborhoods now defines an open cover of the set $A$, regardless of $x$. Recall that $A$ was compact, hence it has a finite subcover of open sets. That is, there exists a finite subset $A' \subset A$, so that the family $\{M_{(x,y)} | \, y \in A'\}$ is also a cover of $A$. Now let $N = \bigcap_{y \in A'} N_{(x,y)}$, then $N$ is open as it is the intersection of finitely many open sets. Furthermore, it is easily seen that $N \cap A = \varnothing$, since $N \cap \left[ \bigcup_{y \in A'} M_{(x,y)} \right] = \varnothing$ and $A \subset \bigcup_{y \in A'} M_{(x,y)}$. Trivially, also $x \in N$. Thus $N$ is an open neighborhood of $x$. Hence, $A^c \subset \text{int } A$, implying that $A^c$ is open. Therefore $A$ is a closed set. □

To conclude, we state Cantor's intersection theorem.

**Lemma 2.23** (Cantor's intersection theorem)**.** *If $E$ is Hausdorff and $A_n$ a decreasing sequence of non-empty compact subsets of $E$, then*

$$\bigcap_{n=0}^{\infty} A_n \neq \varnothing.$$

*Proof.* Let us argue by contradiction, thus suppose that for a decreasing sequence $A_n$ of non-empty compact subsets of a Hausdorff space $E$, the intersection $\bigcap\limits_{n=0}^{\infty} A_n$ is empty. Define the sets $B_n = A_0 \setminus A_n$, then

$$\bigcup_{n=0}^{\infty} B_n = A_0 \setminus \bigcup_{n=0}^{\infty} A_n = A_0 \setminus \varnothing = A_0$$

The sets $B_n$ are also open in $A_0$ as they are complements of closed sets ($A_n$ are closed due to Lemma 2.22). Thus the sets $B_n$ are an open cover of $A_0$. Since $A_0$ was compact, there must exist a finite subcover $\{B_{k_0}, B_{k_1}, \ldots, B_{k_m}\}$ of $A_0$, consisting of $B_n$ sets. It is easily seen that $B_n$ is an increasing sequence of sets. Hence $B_M = B_{k_m}$, for $M = k_m$, contains all other sets in the finite subcover and $A_M = B_M$. Therefore, $A_M = A_0 \setminus B_M = B_M \setminus B_M = \varnothing$, thus a contradiction implying that the intersection $\bigcap\limits_{n=0}^{\infty} A_n$ must be non-empty. □

# 3 The Krein-Milman theorem and the integral version

The main topic of this thesis is the Krein-Milman theorem named after the mathematicians Mark Krein and David Milman, who where the first to prove the statement in [2]. This theorem demonstrates the relation between a compact convex set and its extreme points. More specifically, only information of the set of extreme points allows us to construct each element in the set. In this section we first present a proof of the Krein-Milman theorem, which is largely due to [10]. Afterwards, the integral version of the Krein-Milman theorem is proved as well. Throughout this section, let $E$ be a topological vector space over the field of real numbers.

## 3.1 The Krein-Milman theorem

Here we provide a proof of the Krein-Milman theorem. In the proof of the theorem the next lemma will contribute nicely.

**Lemma 3.1.** *Let $A$ be a nonempty compact subset of a real locally convex space $E$. For any continuous linear functional $f$ on $E$ let*

$$\alpha = \sup_{x \in A} f(x)$$

*Then the set $B = \{x \in A : f(x) = \alpha\}$ is an extreme subset of $A$.*

*Proof.* Firstly, the set $B$ is nonempty, which follows from the extreme value theorem. Hence also the supremum becomes a maximum. Now let $x_1, x_2 \in A$ and $\lambda \in (0,1)$ so that $\lambda\, x_1 + (1-\lambda)\, x_2 \in B$. Let us show that then also $x_1, x_2 \in B$. Observe that, since $f$ is linear, it is true that

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) = \lambda f(x_1) + (1-\lambda) f(x_2) = \alpha$$

Recall that $\alpha$ is the maximum value attained by $f$ for the elements in $A$. Thus indeed $f(x_1), f(x_2) \leq \alpha$. Suppose $f(x_1) < \alpha$ or $f(x_2) < \alpha$. Then

$$\lambda f(x_1) + (1-\lambda) f(x_2) < \lambda\alpha + (1-\lambda)\alpha < \alpha$$

Which is a contradiction as equality was required. Thus $f(x_1) = f(x_2) = \alpha$ implying that $x_1, x_2 \in B$. $\square$

Let us proceed immediately with the proof of the Krein-Milman theorem.

**Theorem 3.2** (Krein-Milman)**.** *Let $E$ be locally convex and Hausdorff. If $A \subset E$ is nonempty, compact and convex, then* ext $A \neq \varnothing$ *and* $\overline{\text{conv}}$ ext $A = A$.

*Proof.* Firstly, let us show that $A$ has extreme points. Let $\mathcal{X}$ be the family of nonempty compact extreme subsets of $A$ and equip this set with the partial ordering defined by $X \leq Y \iff Y \subset X$. Here $\mathcal{X} \neq \varnothing$ since $A \in \mathcal{X}$. Now choose $\mathcal{C} \subset \mathcal{X}$ to be any chain within $\mathcal{X}$. Observe that the intersection of all elements in this chain $\bigcap_{C \in \mathcal{C}} C$ is non-empty due to Cantor's intersection theorem. Furthermore, recall that any intersection of faces is once again a face, thus $\bigcap_{C \in \mathcal{C}} C$ is a face of $A$ . Consequently, by the

Kuratowski-Zorn lemma, the family $\mathcal{X}$ has maximal elements.

Choose a maximal element $X \in \mathcal{X}$. Proceeding by contradiction, assume that there exist $a, b \in X$, with $a \neq b$. Then $a$ and $b$ may be strictly separated by the Hahn-Banach theorem, that is, there exists a continuous linear functional $f$ so that $f(a) < f(b)$. Setting $\gamma = \max_{x \in X} f(x)$ and $X' = \{x \in X : f(x) = \gamma\}$, it follows from Lemma 3.1 that $X'$ is an extreme subset of $X$. It is in fact a proper subset as $a \notin X'$. Hence $X \leq X'$, which is a contradiction since $X$ was assumed to be a maximal element of the family $\mathcal{X}$. Hence $X$ can only contain one element. By definition, if an extreme subset of $A$ contains only one element, then this element is an extreme point of $A$. Thus $A$ has at least one extreme point.

Now follows the inclusion $\overline{\text{conv}}$ ext $A \subset A$. Observe that since $E$ is Hausdorff, $A$ must be closed as it is compact. Hence $\overline{\text{conv}}\ A = A$, as $A$ was assumed to be convex also. Trivially, ext $A \subset A$ by definition of extreme points. Now observe that $A$ is a closed convex set containing ext $A$. Thus $A$ is an element of the family of closed convex sets containing ext $A$. Implying that certainly $\overline{\text{conv}}$ ext $A \subset A$.

All that remains is the backward inclusion. Here, let $C$ denote the closed convex hull of the extreme points of $A$. Let us proceed by contradiction, thus suppose that $A \not\subset C$. Fix any $a \in A$ so that $a \notin C$. Observe that due to Hahn-Banach, and since $C$ is closed and convex, there exists a linear continuous functional $f$ on $E$ such that $f(a) > \alpha$, where $\alpha = \sup_{x \in C} f(x)$. Then the set $S_1 = \{x \in C : f(x) = \alpha\}$ is an extreme subset of $C$. This follows from Lemma 3.1. Furthermore, let $\beta = \sup_{x \in A} f(x)$. Then $S_2 = \{x \in A : f(x) = \beta\}$, defines an extreme subset of $A$. Subsequently, one can see that $\beta > \alpha$ as $f(a) > \alpha$ and $a \in A \setminus C$. That is, for all elements $x \in C$ it is true that $f(x) \leq \alpha < \beta$. This implies that $S_2$ and $C$ must be disjoint. Thus $A$ has an extreme subset disjoint from $C$. Therefore there exists an extreme point in $A \setminus C$, namely within $S_2$. This cannot be true as it must be contained within ext $A$ and thus $C$. Hence a contradiction, thus $A \subset C$. $\square$

## 3.2 Integral representation of the Krein-Milman theorem

This section establishes the implication of the integral form of Krein-Milman from the classical statement. The proof that will be presented here is largely due to Phelps [5]. Firstly, let us introduce the concept of representation by measures.

**Definition 3.3.** *Let $E$ be a locally convex space, $A \subset E$ nonempty and compact and $\mu$ a probability measure on $A$. A point $x \in E$ is represented by the measure $\mu$ if $f(x) = \int_A f \, d\mu$ for all continuous linear functionals $f$ on $E$.*

The integral form of the Krein-Milman theorem gives a criterion for when a point can be represented by a measure.

**Theorem 3.4** (Integral representation of Krein-Milman)**.** *Let $A$ be a non-empty compact and convex subset of a locally convex Hausdorff space $E$. Then every point in $x \in A$ can be represented by a probability measure $\mu$ on $\overline{\text{ext}}\, A$. That is*

$$f(x) = \int\limits_{\overline{\text{ext}}\, A} f(x) \, d\mu(x)$$

*for every continuous linear functional $f$ on $E$.*

*Proof.* As a consequence of the Krein-Milman theorem, any $x \in A$ may be represented as the limit of a net $(x_\tau)_{\tau \in T}$ of the form $x_\tau = \sum_{i=1}^{n_\tau} \lambda_i^\tau \, y_i^\tau$, where $y_i^\tau$ are extreme points of $A$, $\lambda_i^\tau > 0$ and $\sum_{i=1}^{n_\tau} \lambda_i^\tau = 1$. Let $\varepsilon_i^\tau$ be the probability measure on $A$ defined by $\varepsilon_i^\tau(B) = 1$, if $y_i^\tau \in B$, and zero if $y_i^\tau \notin B$. Then the measure $\mu_\tau = \sum \lambda_i^\tau \, \varepsilon_i^\tau$ is a discrete probability measure and represents the point $x_\tau$ since for any continuous linear functional $f$ on $E$ it follows that

$$f(x_\tau) = \sum_{i=1}^{n_\tau} \lambda_i^\tau \, f(x_i^\tau) = \int_{\overline{\text{ext}}\, A} f \, d\mu_\tau .$$

Let $\mathcal{M}$ be the set of all regular Borel measures on $E$ and endow it with the weak*-topology. Furthermore, let $\mathcal{P} \subset \mathcal{M}$ be the set of all probability measures on $\overline{\text{ext}}\, A$. Then, due to the Riesz-Markov-Kakutani representation theorem, there exists a mapping $\rho : \mathcal{P} \to C(\overline{\text{ext}}\, A)^*$ defined by $(\, \rho(\mu) \,)(f) = \int_{\overline{\text{ext}}\, A} f \, d\mu$, which is a homeomorphism onto its image.

Let $\mathcal{Q}$ be the image of $\rho$, then it is easily seen that $\mathcal{Q} = \{ f \in C(\overline{\text{ext}}\, A)^* \, : \, f \geq 0, \;\, f(\mathbb{1}) = 1 \}$. Therefore, $\mathcal{Q}$ is a subset of the unit ball $B_1$ in $C(\overline{\text{ext}}\, A)^*$, which is compact with respect to the weak*-topology in $C(\overline{\text{ext}}\, A)^*$ by the Banach-Alaoglu theorem. Let us show that $\mathcal{Q}$ is closed. Let $\varphi$ be an accumulation point of $\mathcal{Q}$ and $(\varphi_\tau)_{\tau \in T}$ a net converging to $\varphi$ with respect to the weak*-topology. Then, by pointwise convergence of the weak*-topology, it follows easily that $\varphi(\mathbb{1}) = \lim_{\tau \in} \varphi_\tau(\mathbb{1}) = 1$ and $\varphi(f) = \lim_{\tau \in T} \varphi_\tau(f) \geq 0$ for all $f \in C(\overline{\text{ext}}\, A)^*$, implying that $\varphi \in \mathcal{Q}$. So $\mathcal{Q}$ is a closed subset of a compact set, hence $\mathcal{Q}$ must be compact also.

It follows from the continuity of $\rho^{-1}$ that the set $\mathcal{P}$ is compact in $\mathcal{M}$. Consequently, $\mu_\tau$

is a net in the compact set $\mathcal{P}$, which implies that it has a subnet $(\mu_\kappa)_{\kappa \in K}$ converging to some probability measure $\mu$ with respect to the weak*-topology. Observe that, since the net $x_\tau$ converges to $x$, the subnet $x_\kappa$ converges to $x$ also. Now for any bounded linear functional $f$ on $E$, it follows that

$$f(x) = f(\lim_\kappa x_\kappa) = \lim_\kappa f(x_\kappa) = \lim_\kappa \int_A f \, d\mu_\kappa = \int_A f \, d\mu.$$

Thus $x$ is represented by the measure $\mu$. Clearly, $\mu$ is a probability measure on $A$ that is supported on the $\overline{\text{ext}}\, A$, which concludes the proof. $\square$

# 4 Applications of the Krein-Milman theorem

We will now proceed with various is applications of the Krein-Milman theorem. First is a proof of the Stone-Weierstrass theorem.

## 4.1 The Stone-Weierstrass theorem

A well known result in functional analysis is the Stone-Weierstrass theorem, due to Stone in [13]. As a generalization of the Weierstrass approximation theorem, the Stone-Weierstrass theorem admits a criterion for a vector subspace of continuous complex valued functions to be dense within this space. This section provides a proof of this result using the Krein-Milman theorem, which is largely inspired by the proof of de Branges in [3]. Throughout this section $E$ is a topological vector space over the complex numbers.

Let us start with some terminology and notation. For a vector subspace $S$ of $C_0(E,\mathbb{C})$, let $M(S)$ be the set of all real Borel measures $\mu$ on $E$ that have the following two properties: $\mu$ has a total variation of at most 1 and for any $f \in S$ it follows that $\int f\,d\mu = 0$. Observe that $M(S)$ is non-empty as the zero-measure is always contained, regardless of the subspace $S$. Equipping $M(S)$ with the weak*-topology turns it into a locally convex Hausdorff space. Some basic properties of $M(S)$ are established in the next lemma.

**Lemma 4.1.** *The set $M(S)$ is compact and convex. Furthermore, if $S$ is not dense in $C_0(E,\mathbb{C})$, then $M(S)$ contains a nonzero element.*

*Proof.* Firstly, let us consider the compactness. Here the dual space $C_0(E,\mathbb{C})^*$ is considered with the weak*-topology and the operator norm. By the Riesz-Markov-Kakutani representation theorem, there exists a mapping $\rho : M(S) \to C_0(E,\mathbb{C})^*$ defined by $\rho(\mu)(f) = \int_E f\,d\mu$, which is a homeomorphism onto its image. Let $\mathcal{Q}$ be the image of $M(S)$ under $\rho$. Now for every measure $\mu \in M(S)$ it follows by the Hahn-Jordan decomposition and the triangle inequality that

$$\left|\int_E f\,d\mu\right| = \left|\int_E (f^+ - f^-)\,d(\mu^+ - \mu^-)\right| \leq \int_E (f^+ + f^-)\,d(\mu^+ + \mu^-) = \int_E |f|\,|d\mu|$$

Consequently, if $f$ is bounded by 1 it follows by the bound on the total variation of $\mu$ that

$$\left|\int_E f\,d\mu\right| \leq \int_E |d\mu| = |\mu|(E) \leq 1.$$

Using this inequality it is easily derived that

$$\Big|\Big|\int_E f\,d\mu\Big|\Big|_{\text{op}} = \sup\left\{\left|\int_E f\,d\mu\right| \,:\, f \in C_0(E,\mathbb{C}) \text{ and } |\,f\,| \leq 1\right\} \leq 1$$

Therefore, the image $\mathcal{Q}$ is a subset of the closed unit ball in $C_0(E,\mathbb{C})^*$, which is compact with respect to the weak*-topology as a consequence of the Banach-Alaoglu theorem. Additionally, $\mathcal{Q}$ is closed in $C_0(E,\mathbb{C})^*$ by pointwise convergence. Hence $\mathcal{Q}$

is a closed subset of a compact set, implying that it is compact as well. Recall that $\rho$ is a homeomorphism. Therefore, $M(S)$ is compact with respect to the weak*-topology.

Now consider convexity, let $\mu_1, \mu_2 \in M(S)$ and $\lambda \in (0,1)$. Defining $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$, observe that

$$||\,\mu\,|| \leq \lambda\,||\,\mu_1\,|| + (1-\lambda)\,||\,\mu_2\,|| \leq 1.$$

Thus the total variation of $\mu$ is at most 1. Now for any $f \in S$ it follows by linearity in the measures that

$$\int_E f\,d\mu = \lambda \int_E f\,d\mu_1 + (1-\lambda)\int_E f\,d\mu_2 = 0.$$

Therefore, $\mu \in M(S)$ implying that $M(S)$ is convex. Now suppose that $S$ is not dense in $C_0(E,\mathbb{C})$. Then, since $S$ is not dense in $C_0(E,\mathbb{C})$, there exists a function $f$ in the complement of $\overline{S}$. This function can be strictly separated from $\overline{S}$ as a consequence of the Hahn-Banach theorem. That is, there exists a functional $\varphi \in C_0(E,\mathbb{C})^*$ satisfying $|\,\mathrm{Re}(\,\varphi(f)\,)\,| > \sup_{g\in\overline{S}} |\,\mathrm{Re}(\,\varphi(g)\,)\,|$. Observe that it is required that $\varphi(g) = 0$ for all $g \in S$. To see this, suppose this were not true. Then, by linearity and the fact that $S$ is a vector subspace of $C_0(E,\mathbb{C})$, it follows that $\sup_{g\in\overline{S}} |\,\mathrm{Re}(\,\varphi(g)\,)\,| = \infty$. This contradicts the strict separation of $f$ and $\overline{S}$.

By the Riesz-Markov-Kakutani representation, it follows that $\varphi(f) = \int f\,d\mu$, for some measure $\mu$. Then $\mu$ is a nonzero measure and satisfies $\int g\,d\mu = 0$, for all $g \in S$. Furthermore, $\mu$ can be scaled according to $||\,\mu\,|| \leq 1$. If $\mu$ is a complex valued measure, then $\mathrm{Re}(\mu)$ is a real valued measure satisfying the same properties. Therefore, $M(S)$ contains a nonzero element. □

**Lemma 4.2.** *Let $S \subset C_0(E,\mathbb{C})$ be closed under complex conjugation and $\mu$ a nonzero extreme point of the space $M(S)$. If $f$ is a complex valued Borel measurable function on $E$ that is essentially bounded with respect to $\mu$ and satisfies*

$$\int_E fg\,d\mu = 0, \quad \textit{for all } g \in S,$$

*then $f$ is a constant almost everywhere with respect to $\mu$.*

*Proof.* Let $\mu$ be a non-zero extreme point of the space $M(S)$. Observe that such a point exists, otherwise the set $M(S)$ would consist only of the zero measure as a consequence of the Krein-Milman theorem. Observe that $|\mu|(E) \in [0,1]$ since the total variation is bounded from below by 0, and from above by 1. If $|\mu|(E) = 0$, then it follows that $\mu$ must equal the zero measure, leading to a contradiction. If $|\mu|(E) \in (0,1)$, then there exists a $\lambda > 0$ such that $|\mu| < |\lambda\mu|$ and the total variation measure $|\lambda\mu|$ is bounded from above by 1. Observe that also $\int f\,d(\lambda\mu) = 0$, for all $f \in S$, hence $\lambda\mu \in M(S)$. Consequently, the measure $\mu$ is a convex combination of $\lambda\mu$ and the zero measure, implying that $\mu$ is not an extreme point. Therefore, it is required that $|\mu|(E) = \int |d\mu| = 1$.

Let us proceed by contradiction. Without loss of generality let $f$ be a positive real valued Borel measurable function satisfying $\int f\,|d\mu| = 1$ and the assumptions of the lemma, but is not equal to a constant almost everywhere. Let $\beta^{-1} > 0$ be the essential bound of $f$ with respect to $\mu$ and suppose that $\beta \geq 1$. Then it follows that

$$\int_E |1-f|\,|d\mu| = \int_E 1-f\,|d\mu| = \int_E |d\mu| - \int_E f\,|d\mu| = 0.$$

Therefore $f = 1$ almost everywhere with respect to $\mu$, which contradicts the hypotheses that $f$ was a constant almost everywhere. Hence it is required that $\beta \in (0,1)$. Consequently, we can define the two Borel measures $\mu_1, \mu_2$ on $E$ by

$$\mu_1(A) = \int_A f\,d\mu, \quad \mu_2(A) = \int_A \frac{1-\beta f}{1-\beta}\,d\mu.$$

Clearly, $\mu$ is a convex combination of $\mu_1$ and $\mu_2$. All that remains is to check that both $\mu_1$ and $\mu_2$ belong to $M(S)$. It is easily seen that the total variation of both $\mu_1$ and $\mu_2$ are bounded from above by 1. This is a consequence of the assumption $\int f\,|d\mu| = 1$ and the property that the measure $\mu$ has a total variation of at most 1. Consequently, by the definition of integrals and $\mu_1$ and $\mu_2$ one can show that for all $g \in S$

$$\int_E g\,d\mu_1 = \int_E fg\,d\mu, \quad \int_E g\,d\mu_2 = \int_E g\frac{1-\beta f}{1-\beta}\,d\mu.$$

Both of these integrals evaluate to zero by the assumptions on $f$, Therefore, the second property is also satisfied, implying that $\mu_1, \mu_2 \in M(S)$. Hence $\mu$ is no extreme point of $M(S)$, contradicting the assumptions and thus proving the statement. □

**Theorem 4.3** (Stone-Weierstrass)**.** *Let $E$ be locally compact and Hausdorff. If $S$ is a vector subspace of $C_0(E,\mathbb{C})$ that satisfies the following properties:*

(i) *For all $a \in E$, there exists a function $f \in S$ such that $f(a) \neq 0$;*

(ii) *For all $a, b \in E$, $a \neq b$, there exists a function $f \in S$ such that $f(a) \neq f(b)$;*

(iii) *If $f, g \in S$, then $fg \in S$;*

(iv) *If $f \in S$, then $\bar{f} \in S$ .*

*Then the set $S$ is dense in the space $C_0(E,\mathbb{C})$.*

*Proof.* Let $S$ be a subspace of $C_0(E,\mathbb{C})$ that satisfies the properties (ii)-(iv) and is not dense within $C_0(E,\mathbb{C})$. Let us derive a contradiction to property (i). Due to the Lemmas 4.1 and 4.2, there exists a nonzero extreme point $\mu$ of $M(S)$. Let $f \in S$ be an arbitrary function. Then for any $g \in E$, the product $fg$ is also in $E$ by property (iii). Since $\mu \in M(S)$, it follows that $\int fg\,d\mu = 0$. Then Lemma 4.2 implies that $f = \lambda_f$ almost everywhere with respect to $\mu$, where $\lambda_f$ is some constant.

By continuity of $f$, it follows that $f = \lambda_f$ on the support of $\mu$. Additionally, the support of $\mu$ consists of only one point. To see this take any two points $a, b \in S$, such that $a \neq b$. Then by (ii) there must exist a $g$ that separates these points. However, $g(a) = g(b)$ on the support of $\mu$ as $g$ was constant here. Therefore, it is required that $a = b$, implying that supp $\mu = \{ p \}$, for some $p \in S$.

Recall from the proof of Lemma 4.2 that $|\mu|(E) = 1$ and therefore $|\mu|(\{ p \}) = 1$, as $\mu$ is an extreme point of $M(S)$ and the support of $\mu$ consists only of the point $p$. Observe that $|\mu|(\{ p \}) = | \mu(\{ p \}) |$ since the only subset of $\{p\}$ different from itself is the empty set. Therefore, $\mu(\{ p \}) = 1$. Now the equality $\int f \, d\mu = 0$ implies that $\lambda_f \, \mu(\{ p \}) = 0$, hence $\lambda_f = 0$. Thus $f(p) = 0$, contradicting property (i) since $f$ was chosen arbitrarily. $\square$

## 4.2 Completely monotone functions and Berstein's theorem

This section provides an application of the integral Krein-Milman theorem on the class of completely monotone functions. Using this theorem one can prove Bernstein's theorem, which says that any completely monotone function may be constructed from exponentials. More specifically, it states that such a function is the Laplace transform of a non-negative finite Borel measure. The proof of Bernstein's theorem presented here is largely inspired by [4] and [5]. Recall that a function $f : (0, \infty) \to [0, \infty)$ is called *completely monotone* if it has continuous derivatives of all orders and satisfies the following property

$$(-1)^n f^{(n)}(t) \geq 0, \quad \text{for } t > 0 \text{ and } n = 0, 1, 2, \ldots$$

It follows from the definition that completely monotone functions are non-negative ($f(t) \geq 0$), non-increasing ($f'(t) \leq 0$) and convex ($f''(t) \geq 0$). Furthermore, if $f$ and $g$ are completely monotone, then the product and any linear combination of the two is also completely monotone.

Let $C^{\infty}(0, \infty)$ be the set of infinitely differentiable continuous real valued functions on the interval $(0, \infty)$. One can check that the functions

$$p_n(f) = \max_{t \in [\frac{1}{n}, n]} |f^{(n-1)}(t)|, \quad n \in \mathbb{N} \tag{1}$$

are semi-norms on $C^{\infty}(0, \infty)$. Equipping $C^{\infty}(0, \infty)$ with the topology generated by the family of these semi-norms turns it into a locally convex Hausdorff space. The Hausdorff property is satisfied since the family of semi-norms is separated. Recall that a topological vector space is called *Montel* if any bounded and closed subset is also compact. As demonstrated in [4] and [14], the space $C^{\infty}(0, \infty)$ is Montel. Observe that the family of completely monotone functions is contained in $C^{\infty}(0, \infty)$. Let us study the family $\mathcal{M}$ of completely monotone functions that are bounded from above by 1. One can deduce interesting results regarding this family. However, before considering these, let us present a convenient lemma regarding convergence that will be used onwards.

**Lemma 4.4.** *If a net $f_\tau \to f$ in $C^{\infty}(0, \infty)$, then $f_\tau \to f$ pointwise.*

*Proof.* Without loss of generality, suppose $(f_\tau)_{\tau \in T}$ is a net converging to zero. For any fixed $n \in \mathbb{N}$, let us show that $f_\tau$ converges pointwise on the interval $[\frac{1}{n}, n]$. To achieve this, we will derive that $f_\tau^{(n-k)}$ converges pointwise on $[\frac{1}{n}, n]$, for $1 \leq k \leq n$. The proof will be by induction on $k$. Consider first the case where $k = 1$. Then, by Proposition 2.8 it is easily seen that

$$p_n(f_\tau) = \max_{t \in [\frac{1}{n}, n]} |\, f_\tau^{(n-1)}(t) \,| \longrightarrow 0.$$

Therefore, $f_\tau^{(n-1)}$ converges pointwise on $[\frac{1}{n}, n]$. Proceeding, suppose that $f_\tau^{(n-k)}$ converges pointwise on $[\frac{1}{n}, n]$, for some $1 \leq k < n$. Then, by the fundamental theorem of calculus, it follows that

$$f_\tau^{(n-k-1)}(t) = f_\tau^{(n-k)}\left(\frac{1}{n}\right) + \int_{\frac{1}{n}}^{t} f_\tau^{(n-k)}(x)\, \mathrm{d}x \longrightarrow 0.$$

Thus $f_\tau^{(n-k)}$ converges pointwise on $[\frac{1}{n}, n]$, for $1 \leq k \leq n$, concluding the inductive part of the proof. Consequently, $f_\tau$ converges pointwise on the interval $[\frac{1}{n}, n]$. Observe that for arbitrary $t \in (0, \infty)$, there exists an $m \in \mathbb{N}$ such that $t \in [\frac{1}{m}, m]$. Hence $f_\tau$ converges pointwise on $(0, \infty)$. $\square$

The next lemma deals with the criteria necessary for the application of the Krein-Milman theorem.

**Lemma 4.5.** *The family $\mathcal{M} \subset C^\infty(0, \infty)$ is convex and compact.*

*Proof.* Let $f_1, f_2 \in \mathcal{M}$ and $\lambda \in (0, 1)$. Take $f = \lambda f_1 + (1 - \lambda) f_2$, then

$$(-1)^n f^{(n)}(t) = \lambda (-1)^n f_1^{(n)}(t) + (1 - \lambda)(-1)^n f_2^{(n)}(t) \geq 0$$

for all $t > 0$ since $f_1, f_2 \in \mathcal{M}$. Thus $f$ is completely monotone. Furthermore,

$$f(t) = \lambda f_1(t) + (1 - \lambda) f_2(t) \leq \lambda + (1 - \lambda) = 1$$

for all $t > 0$ implying that $f$ is also bounded by 1. Thus $f \in \mathcal{M}$. Hence $\mathcal{M}$ is convex in $C^\infty(0, \infty)$.

Since $C^\infty(0, \infty)$ is a Montel space, compactness of $\mathcal{M}$ is equivalent to it being closed and bounded. Let $f$ be an accumulation point of $\mathcal{M}$, then there exist a net $(f_\tau)_{\tau \in T}$ in $\mathcal{M}$ converging to $f$. Additionally, the mapping of taking first order derivatives is continuous. To see this, observe that for arbitrary $n \in \mathbb{N}$ and all $g \in C^\infty(0, \infty)$ it follows that

$$p_n(\, g' \,) = \max_{t \in [\frac{1}{n}, n]} |\, g^{(n)}(t) \,| \leq \max_{t \in [\frac{1}{n+1}, n+1]} |\, g^{(n)}(t) \,| = p_{n+1}(\, g \,).$$

Therefore, by Proposition 2.6, the mapping of taking first order derivatives is continuous. Consequently, by induction on $n$, the mapping of taking $n$-th order derivatives is continuous as well. Thus, by continuity of the differential operator

$$(-1)^n \frac{\mathrm{d}^n f}{\mathrm{d}t^n} = (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}t^n} \lim_{\tau \in T} f_\tau(t) = \lim_{\tau \in T} (-1)^n \frac{\mathrm{d}^n f_\tau}{\mathrm{d}t^n} \geq 0.$$

Now together with the pointwise convergence established in Lemma 4.4, we conclude that the limit $f$ is also completely monotone. One can show that $f$ is also bounded by the same argument. Therefore, $f \in \mathcal{M}$ implying that $\mathcal{M}$ contains all its accumulation points. Hence $\mathcal{M}$ is closed.

Let us proceed with showing that $\mathcal{M}$ is bounded. One can conclude that $\mathcal{M}$ is bounded whenever $\sup\{p_n(f) : f \in \mathcal{M}\}$ is finite for all $n \in \mathbb{N}$. This is exactly Proposition 2.7. Observe that this is implied whenever $\sup_{t \in [a, \infty)} |f^{(n)}(t)|$ is finite for all $n \in \mathbb{N}$ and $a \in (0, 1)$. Proceeding by induction on $n$, let us show that for any function $f \in \mathcal{M}$

$$|f^{(n)}(t)| \leq a^{-n}\, 2^{(n+1)(n/2)}, \quad n \in \mathbb{N},\ t \in [a, \infty).$$

The base case is trivial since all functions in $\mathcal{M}$ are required to be bounded from above by 1, and from below by 0. Now assume that the induction hypothesis holds in the case

$n$. Applying the mean value theorem to the function $f^{(n)}$ on the interval $[\frac{a}{2}, a]$ implies the existence of a constant $c \in [\frac{a}{2}, a]$ such that $(\frac{a}{2})\, f^{(n+1)}(c) = f^{(n)}(a) - f^{(n)}(\frac{a}{2})$. From this it is easily derived that $f^{(n)}(\frac{a}{2}) \geq -(\frac{a}{2})\, f^{(n+1)}(c)$ which implies that

$$(-1)^n f^{(n)}\left(\frac{a}{2}\right) \geq (-1)^{n+1}\left(\frac{a}{2}\right) f^{(n+1)}(c) \geq (-1)^{n+1}\left(\frac{a}{2}\right) f^{(n+1)}(a).$$

Since either both $(-1)^n$ and $f^{(n)}$ are non-negative or non-positive and $(-1)^{n+1} f^{(n+1)}$ is non-increasing. Hence also

$$\left|f^{(n+1)}(a)\right| = (-1)^{n+1} f^{(n+1)}(a) \leq (-1)^n \left(\frac{2}{a}\right) f^{(n)}\left(\frac{a}{2}\right) = \left(\frac{2}{a}\right)\left|f^{(n)}\left(\frac{a}{2}\right)\right|.$$

Now applying the induction hypothesis at $\frac{a}{2}$ gives that

$$\left|f^{(n+1)}(a)\right| \leq \left(\frac{2}{a}\right)\left(\frac{a}{2}\right)^{-n} 2^{(n+1)(n/2)} = a^{-(n+1)}\, 2^{\frac{1}{2}(n+1)(n+2)}$$

Therefore, the upper bound we set out to show is valid at $a$ and thus also at any other value by the non-increasing property of $|f^{(n+1)}|$. Hence $\sup_{t\in[a,\infty)} |f^{(n)}(t)|$ is finite for all $n \in \mathbb{N}$. Consequently, the set $\mathcal{M}$ is bounded. Thus $\mathcal{M}$ is compact in $C^{\infty}(0, \infty)$. □

Since the set $\mathcal{M}$ is compact and convex in a locally convex Hausdorff space, it contains extreme points as a consequence of the Krein-Milman theorem. A natural question to ask is what completely monotone functions make up the set of extreme points. This question is answered by the next statement.

**Lemma 4.6.** *The extreme points of the family $\mathcal{M}$ are the exactly functions $f(t) = a^t$ for $a \in [0, 1]$.*

*Proof.* Firstly, let us show that any extreme point is of this form. Let $f \in \operatorname{ext} \mathcal{M}$ and define, for some fixed $\tau > 0$, the function $\varphi(t) = f(t+\tau) - f(t)f(\tau)$. Now define the functions $f_1 = f - \varphi$ and $f_2 = f + \varphi$ it is easily checked that both $f_1, f_2 \in \mathcal{M}$. Now suppose that $\varphi$ is not the zero function. Then $f$ can be written as $\frac{1}{2}(f_1 + f_2)$ implying that $f$ is not an extreme point. Hence $\varphi$ must be the zero function. Therefore $f$ satisfies $f(t+\tau) = f(t)f(\tau)$. it follows that $f$ must be of the form $a^t$. This is a well-known result in analysis. It is required that $f(t) = a^t$ must be bounded by 1 and completely monotone. To satisfy complete monotonicity it is required that

$$(-1)^n \frac{\mathrm{d}^n a^t}{\mathrm{d}t^n} \geq 0 \Longrightarrow (-1)^n \ln(a)^n a^t \geq 0$$

which is satisfied exactly when $a \in [0, 1]$. At the same time the requirement to be bounded by 1 is also satisfied.

It remains to show that any function of the form $a^t$ for $a \in [0, 1]$ is also an extreme point. Trivially, the constant functions 0 and 1 are extreme points of $\mathcal{M}$. This is due to the fact that functions in $\mathcal{M}$ are bounded from above by 1, and by 0 from below.

If these were the only extreme points then any function would be constant due to the Krein-Milman theorem. Thus there must exist at least one other extreme point, which is an exponential function. Let $\varepsilon(t) = \alpha^t$ be this extreme point, where $\alpha \in (0,1)$. Now define the bijective mapping $B : \mathcal{M} \to \mathcal{M}$ by $B(f) = f(\beta t)$ for some arbitrary $\beta \geq 0$. This mapping preserves convex combinations since

$$B(\,\lambda f_1 + (1-\lambda) f_2\,) = \lambda f_1(\beta t) + (1-\lambda) f_2(\beta t)$$

for any $f_1, f_2 \in \mathcal{M}$ and $\lambda \in [0,1]$. Therefore an extreme point gets mapped onto an extreme point. Hence the function $B(\varepsilon)$ is also an extreme point. Since $\beta$ was arbitrarily chosen, any function $f(t) = a^t$ for $a \in [0,1]$ is an extreme point, which concludes the proof. $\square$

Using the information about the extreme points of $\mathcal{M}$ one can proceed to prove Bernstein's theorem using the integral Krein-Milman theorem. The statement was first proved by Bernstein in [15].

**Theorem 4.7** (Bernstein)**.** *Let $f$ be a bounded and completely monotonic function, then there exists a probability measure $\mu$ on the interval $[0,1]$ such that for all $t > 0$*

$$f(t) = \int\limits_{[0,1]} a^t \, d\mu(a)$$

*Proof.* Again, let $\mathcal{M}$ be the family of completely monotone functions bounded by 1. Due to Lemma 4.5 and 4.6 it is clear that the extreme points of $\mathcal{M}$ are the exponential functions $a^t$ for $a \in [0,1]$. The set of extreme points is also closed, this follows from Lemma 4.6. It follows now from the Krein-Milman theorem in integral form that there exists a probability measure $\nu$ on ext $\mathcal{M}$ such that

$$f = \int\limits_{\text{ext}\,\mathcal{M}} I \, d\nu$$

for any $f \in \mathcal{M}$. Here $I$ denotes the identity mapping. Now define the function $M : [0,1] \to \text{ext}\,\mathcal{M}$ defined by $M(a) = a^t$. Then $M$ defines a continuous function. Furthermore, it's inverse exists and is continuous also, thus $M$ is a homeomorphism. Define now the measure $\mu$ on the interval $[0,1]$ by $\mu(A) = \nu(\,M(A)\,)$. Thus $\mu$ measures the image of a set under the mapping $M$. Therefore the integral above may be rewritten as

$$f = \int\limits_{\text{ext}\,\mathcal{M}} I \, d\nu = \int\limits_{[0,1]} M(a) \, d\mu(a)$$

Implying that

$$f(t) = \int\limits_{[0,1]} a^t \, d\mu(a)$$

So far we have only addressed completely monotonic functions that are bounded by 1. For general completely monotone functions the same result is achieved via normalization of the function. This concludes the proof. $\square$

## 4.3 Doubly stochastic matrices and Kendall's theorem

This section provides a proof of Kendall's theorem using the Krein-Milman theorem. It is named after the mathematician David Kendall who first proved the result in [16]. As a generalization of the Birhkoff-von Neumann theorem, this theorem states that infinite doubly stochastic matrices may be approximated by convex combinations of permutation matrices.

Let us first introduce some definitions. A finite or infinite square matrix is called *doubly stochastic* if all of its entries are non-negative real numbers and the row- and column-sums are all equal to 1. A class of such matrices are the permutation matrices. A *permutation matrix* is a square matrix in which all rows and columns contain exactly one entry that is equal to one, and has all other entries equal zero. Multiplying any matrix with a permutation matrix of equal dimension will permute the order of rows or columns of this matrix. Lastly, for a square matrix $A$ with dimension $n$ we set $\dim A = \{1, 2, \ldots, n\}$. Intuitively, if $A$ is infinite, set $\dim A = \mathbb{N}$.

The setting of this section will be the following. The spaces $M_n(\mathbb{R})$ of finite $n$-dimensional square real valued matrices, and $M_\infty(\mathbb{R})$ infinite square real valued matrices are Banach spaces when equipped with the supremum norm defined by

$$|| A ||_\infty = \sup\{ | a_{ij} | \, : \, (i,j) \in \dim A \times \dim A \}.$$

Consequently, we will write $M(\mathbb{R})$ for these spaces if the distinction between finite and infinite is not necessary. Equipping this space with the topology generated by the supremum norm turns it into a locally convex Hausdorff space. Let us study the set $\mathcal{D} \subset M(\mathbb{R})$ of doubly stochastic matrices, which is convex and compact as demonstrated by the next lemma.

**Lemma 4.8.** *The set $\mathcal{D}$ is compact and convex in $M(\mathbb{R})$.*

*Proof.* Consider first the compactness of $\mathcal{D}$. It is easily seen that the limit of a sequence of doubly stochastic matrices in $M(\mathbb{R})$ is again doubly stochastic. In the finite case it is now easily checked that $\mathcal{D}$ is compact. This is a direct consequence of $\mathcal{D}$ being complete and totally bounded. For the infinite case, observe that these matrices may be viewed as elements of $\mathbb{R}^\infty$ by simply rewriting the entries. Note that there are multiple ways to achieve this, but this will not matter. Let $\varphi : M_\infty(\mathbb{R}) \to \mathbb{R}^\infty$ be the mapping associating each matrix with its representation in $\mathbb{R}^\infty$. Clearly, $\varphi$ is continuous. The image of $\mathcal{D}$ under $\varphi$ takes values in the infinite direct product of the interval $[0, 1]$, which is compact as seen in Section 4.4. Consequently, the image $\varphi(\mathcal{D})$ is compact, since $\varphi(\mathcal{D})$ is closed by continuity. Therefore, $\mathcal{D}$ must be compact also as a consequence of continuity.

Next consider the convexity of $\mathcal{D}$. Let $D_1, D_2 \in \mathcal{D}$ and $\lambda \in (0, 1)$. Setting $D = \lambda D_1 + (1 - \lambda) D_2$ it follows easily that the row- column-sums of $D$ evaluate to 1 and that all entries are non-negative, implying that $D$ is doubly stochastic. Thus $\mathcal{D}$ is a convex set, proving the Lemma. $\square$

Now we can turn our attention towards Kendall's theorem. In the proof of the theorem, the next lemma will contribute greatly, of which the proof is due to Mauldon [6].

**Lemma 4.9.** *If $A$ is a, possibly infinite, real valued nonzero matrix such that $|\, a_{ij} \,| \leq \frac{1}{2}$ for all $i, j$ and that both $\sum_j a_{ij}$ and $\sum_i a_{ij}$ are integers. Then there exists a real valued nonzero matrix $B$ such that $|\, b_{ij} \,| \leq |\, a_{ij} \,|$ for all $i, j$ and whose row- and column-sums are all zero.*

*Proof.* It is easily checked that applying permutations to the rows and columns of the matrix $A$ preserves the assumptions. Let us do so to ensure that $d_{11} \neq 0$. Next we define the notion of a *path* and a *cycle* in a matrix. A path $P_n$ in a matrix $A$ is a finite sequence $(i_n, j_n)$, for $n = 1, 2, \ldots, N$, consisting of distinct indices of the matrix such that

(i) $a_{i_n j_n} \neq 0$ for all $n = 1, 2, \ldots, N$;

(ii) $(i_1, j_1) = (1, 1)$ and either $i_2 = 1$ or $j_2 = 1$;

(iii) if $i_n = i_{n-1}$, then $j_{n+1} = j_n \neq j_{n-1}$, and if $j_n = j_{n-1}$, then $i_{n+1} = i_n \neq i_{n-1}$.

Intuitively, property (iii) ensures that the sequence must swap from row after swapping from column, and vice versa. An endpoint of a path is the index $(i_N, j_N)$. Similar to a path, a cycle is a path starting in an arbitrary non-zero entry, with the first entry equal to the last. Let $P$ denote the collection of all endpoints of all paths in $A$.

For an index $(k, l) \in P$, let us distinct between the cases where there exists either a unique path with this index as its endpoint, or not. Consider first the easy case where the path is not unique. Then there are at least two paths with endpoint $(k, l)$, which need not have disjoint indices. Then there exists a cycle in $A$ which may be constructed from these paths. Let $C_n = (i_n, j_n)$ be this cycle and define the constant

$$b = \min_{(i_n, j_n) \in C_n} |\, a_{i_n j_n} \,|.$$

Consequently, construct the matrix $B$ with indices given by

$$b_{ij} = \begin{cases} 0 & \text{, if } (i,j) \neq C_n \\ (-1)^n\, b & \text{, if } (i,j) = C_n \end{cases}$$

Then it is clear that $|\, b_{ij} \,| \leq |\, a_{ij} \,|$ for all $i$ and $j$. Additionally, the row- and column sums evaluate to zero, since every row and column contains only zero entries or has pairs of two consecutive terms of the cycle (which clearly sum to zero). Thus the matrix $B$ is as required.

Now consider the second case, namely that for every index in $P$ there is a unique path with this index as its endpoint. For any such index $(k, l)$, define the *predecessor* $P(k, l)$ of $(k, l)$ as the previous index on the unique path ending in $(k, l)$. Now observe that $P(k, l)$ and $(k, l)$ either lie in the same row, or in the same column. Regardless, let $\Sigma_{kl}$

be the sum of the entries in this row or column. Consequently, define the coefficients $w_{ij}$ by

$$w_{ij} = \frac{a_{P(i,j)}}{a_{P(i,j)} - \Sigma_{ij}}.$$

Then $|\, w_{ij}\,| \leq 1$, since $|\, a_{ij}\,| \leq \frac{1}{2}$ and $\Sigma_{ij}$ is integer. Using these coefficients, define the coefficients $m_{ij}$ inductively by

$$m_{ij} = \begin{cases} 0 & \text{, if } (i,j) \notin P \\ w_{ij}\, m_{P(i,j)} & \text{, otherwise} \end{cases}$$

At last, we can construct the matrix $B$. Its entries are prescribed by the product $b_{ij} = m_{ij}\, a_{ij}$. All that is left to show is that $B$ is as required. Observe firstly that $|\, m_{ij}\,| \leq |\, m_{P(i,j)}\,|$, since $|\, w_{ij}\,| \leq 1$. From this it follows easily that $|\, m_{ij}\,| \leq 1$, which implies that $|\, b_{ij}\,| = |\, m_{ij}\,|\,|\, a_{ij}\,| \leq |\, a_{ij}\,|$. So that the first condition is satisfied.

Now consider the second property, namely that the row- and columns sums are zero. To this end, fix a nonzero column of $B$, say $c$. Then this row must contain indices from $P$ as it has nonzero entries. Additionally, the column must have exactly two entries. If it had only one, then the row- and column-sums would not evaluate to an integer in $A$ and if there where more than 2, this would contradict the uniqueness of the paths. Let $(k, c)$ and $(l, c)$ be the indices of these entries, with $k < l$. Then it follows that

$$b_{lc} = m_{lc}\, a_{lc} = \frac{a_{kc}}{a_{kc} - \Sigma_{lc}}\, a_{lc}\, m_{kc} = -a_{kc}\, m_{kc} = -b_{kc}.$$

Since the column-sum is simply the sum of the entries $b_{kc}$ and $b_{lc}$, this sum is equal to zero. Therefore, all columns sum to zero. By the same argument it can be shown that the rows of $B$ also sum to zero. Hence $B$ is as required, concluding the proof. $\square$

**Theorem 4.10** (Kendall)**.** *The extreme points of $\mathcal{D}$ are exactly the permutation matrices.*

*Proof.* Firstly, let us show that any permutation matrix is an extreme point. Arguing by contradiction, let $P$ be a permutation matrix and suppose that is not an extreme point. Then $P = \lambda A + (1 - \lambda)B$ with $A, B \in \mathcal{D}$ and $\lambda \in (0, 1)$. Hence if $p_{ij} = 0$, it follows that $\lambda a_{ij} + (1 - \lambda)b_{ij} = 0$, which implies that $a_{ij} = b_{ij} = 0$ since the entries in a doubly stochastic matrix are non-negative. By the same argument, it follows that if $p_{ij} = 1$, then $a_{ij} = b_{ij} = 1$. Therefore, it is required that $A = B = P$, implying that $P$ may only be represented as a convex combination of itself. Thus $P$ is an extreme point.

Now let $D \in \mathcal{D}$ be any doubly stochastic matrix that is not a permutation matrix. Subsequently, construct the matrix $M$ with the entries given by

$$m_{ij} = \begin{cases} d_{ij} & \text{, if } d_{ij} \leq \frac{1}{2} \\ d_{ij} - 1 & \text{, if } d_{ij} > \frac{1}{2}. \end{cases}$$

Observe that if $M$ is the zero matrix, then $A$ must be a permutation matrix, leading to a contradiction. Hence $M$ must be a nonzero matrix. It is obvious that in each row and column of $M$, at most one of the entries $m_{ij}$ has value $a_{ij} - 1$. Therefore, the rows and columns of $M$ must sum to either zero or 1. Additionally, it is true that $|\, m_{ij} \,| \leq \frac{1}{2}$. Thus Lemma 4.9 is applicable to $M$, implying that there exists a real non-zero matrix $B$ which rows and column sum to zero. Then the matrices $A \pm B$ are not equal to $A$ and have row- and column-sums equal to 1. Additionally, these matrices have non-negative entries since $|\, b_{ij} \,| \leq |\, m_{ij} \,| \leq |\, d_{ij} \,|$ for all $i, j$. Therefore, $A$ is a convex combination of $A \pm B$ and thus no extreme point. $\square$

By the Krein-Milman theorem, the set $\mathcal{D}$ is equal to the closed convex hull of the set of permutation matrices. Thus each doubly stochastic matrix may be approximated by convex combinations of permutation matrices. In the finite case, a doubly stochastic matrix may be presented as a (finite) convex combination of permutation matrices, rather than approximated. This is exactly the Birkhoff-von Neumann theorem.

## 4.4 Numbering lattice points.

Consider the integer lattice $\mathbb{Z}^2$ and suppose that we want to associate with each lattice point a real number in the interval $[0,1]$, such that the value assigned to each point is the average of its neighbours. We will show that all such functions are constants between 0 and 1.

Essentially, we are looking for functions $f : \mathbb{Z}^2 \to [0,1]$ such that

$$f(x,y) = \frac{1}{4}\Big[f(x-1,y) + f(x,y+1) + f(x+1,y) + f(x,y-1)\Big],$$

for all $(x,y) \in \mathbb{Z}^2$. This condition will be referred to as the average condition. Recall that the vector space $L_\infty(\mathbb{Z}^2)$ is a Banach space when equipped with the essential supremum norm defined by

$$||\, f\, ||_\infty = \sup\{\, |\, f(x,y)\,| \,:\, (x,y) \in \mathbb{Z}^2\, \},\ f \in L_\infty(\mathbb{Z}^2).$$

It is import to specify that $\mathbb{Z}^2$ is considered as the measure space $(\mathbb{Z}^2, \Sigma,\, \delta)$, where $\Sigma$ is some $\sigma$-algebra on $\mathbb{Z}^2$ and $\delta$ the counting measure. Additionally, equip $L_\infty(\mathbb{Z}^2)$ with the topology of pointwise convergence, which turns it into a locally convex Hausdorff space. Let $\mathcal{F}$ be the set of all non-negative functions satisfying the average condition and $||\, f\, ||_\infty \leq 1$. It is easily checked that $\mathcal{F}$ is a closed set. Furthermore, $\mathcal{F}$ is a subset of $L_\infty(\mathbb{Z}^2)$ that is compact and convex. Indeed, compactness follows from Tikhonov's theorem. To see this, consider $\mathbb{R}^\infty$ with the product-topology and define the *spiral* continuous mapping $\psi : L_\infty(\mathbb{Z}^2) \to \mathbb{R}^\infty$ by

$$\psi(f) = (\, f(0,0),\ f(1,0),\ f(1,1),\ f(0,1),\ \dots),\quad f \in L_\infty(\mathbb{Z}^2).$$

By Tikhonov's theorem, the infinite Cartesian product of the intervals $[0,1]$ is a compact subset of $\mathbb{R}^\infty$, since the interval $[0,1]$ is compact in $\mathbb{R}$ with respect to the usual topology. Observe that the image of $\mathcal{F}$ under $\psi$ is closed due to continuity and is a subset of the infinite direct product of $[0,1]$. Hence, this image is compact also. Thus, by continuity of $\psi$, $\mathcal{F}$ must be compact also.

For convexity, let $f_1, f_2 \in \mathcal{F}$ and take $\lambda \in (0,1)$. Then it is easily checked that the convex combination $f = \lambda f_1 + (1-\lambda) f_2$ satisfies the average condition and the boundedness from below by 0, and from above by 1. Hence $f \in \mathcal{F}$, implying that $\mathcal{F}$ is convex.

Therefore, $\mathcal{F}$ has extreme points as a consequence of the Krein-Milman theorem. Clearly, the constant functions 0 and 1 are extreme points of $\mathcal{F}$. Now suppose that $f$ is an extreme point of $\mathcal{F}$ other than 0 and 1. Since $f$ satisfies the average condition, $f$ is a convex combination of the four shifted functions $f(x-1,y),\ f(x,y+1),\ f(x+1,y),\ f(x,y-1)$. This can only be possible if the shifted functions are identical, since then $f$ would not be an extreme point. Consequently, $f$ is a constant function, leading to a contradiction. Hence the functions 0 and 1 are the only extreme points of $\mathcal{F}$.

By the Krein-Milman theorem, $\mathcal{F}$ is the closure of the convex hull of the set of extreme points. From this it follows trivially that any $f \in \mathcal{F}$ is a constant function between 0 and 1.

## 4.5 Numbering vertices of locally finite graphs

In this section the question asked in Section 4.4 will be generalized from the integer lattice to infinite connected graphs. The integer lattice $\mathbb{Z}^2$ can be considered as an infinite graph, with the coordinates $(x, y) \in \mathbb{Z}^2$ being the vertices of this graph and edges between its neighbours in $\mathbb{Z}^2$. Let us refer to this graph as the Manhattan-graph, which will be denoted by $M$.

An infinite graph $G$ is called *locally finite* if each vertex has a finite degree. The Manhattan-graph is an example of a locally finite graph. For a graph $G$ and a vertex $v \in V(G)$, denote by $N(v)$ the neighbours of $v$ in $G$. The question asked regarding the Manhattan-graph was the following. Do there exist functions $f : V(M) \to [0, 1]$ such that $f(v)$ is the average of the numbers $f(u)$ for $u \in N(v)$, but $f$ is not a constant functions. More generally, for a locally finite graph $G$, this question translates to the following: Are there functions $f : V(G) \to [0, 1]$ such that

$$f(v) = \frac{1}{|N(v)|} \sum_{u \in N(v)} f(u),$$

for all $v \in V(G)$, other than the constant functions. As seen in the previous section, there do not exist such functions for the Manhattan-graph. It turns out, there exist locally finite graphs for which these functions do exist.

We will now present a graph $G$ and construct function $f$ that is not constant but satisfies the average condition. An embedding of a part of the graph $G$ containing the root can be found in Figure 1. To construct the graph, start with a single vertex, which will be referred to as the *root* of $G$. Proceeding recursively, the graph is generated by branching the neighbours of the root into two vertices, and then also the neighbours of these vertices into two, continuing forever.

Let us now construct a function $f$ that satisfies the average condition, but is not constant. Denote by $s$ the root vertex and let $s_1, s_2$ be its neighbours. Define $S_1$ to be the subgraph of $G$ containing $s$ and all vertices $v$ such that $s_1$ lies on every path connecting $v$ and $s$. Similarly, one can define $S_2$. Intuitively, one can view the subgraphs $S_1$ and $S_2$ as the left and right side of the graph $G$. Note that $s$ is in both $S_1$ and $S_2$, but any other vertex is in only one of these. Define the *depth* of a vertex $v$ as the length of the shortest path in $S_1$ or $S_2$ connecting $s$ and $v$, which will be denoted as $\pi(v)$. Observe that by the structure of $S_1$, any path is the shortest path.

The value assigned by $f$ to a vertex will be based on its depth. Define the sequence $(s_n)_{n\geq 0}$ recursively as

$$s_n = \frac{1}{3}\Big(s_{n-1} + 2s_{n+1}\Big),$$

with initial values $s_0 = \frac{1}{2}$ and $s_1 = \frac{1}{2} + \delta$, for some $\delta > 0$. Observe that this is exactly the average condition for all vertices in $S_1$ except the vertex $s$. This is a homogeneous second-order difference equation having constant coefficients, of which the solution can be obtained by finding the roots of the characteristic equation. The roots are $r_1 = 1$ and $r_2 = \frac{1}{2}$, hence the general solution reads $s_n = c_1 + c_2\left(\frac{1}{2}\right)^n$. Applying the initial conditions it follows that

$$s_n = \frac{1}{2} + 2\delta - \delta\left(\frac{1}{2}\right)^{n-1}.$$

By the same construction, one can construct the sequence $(t_n)_{n\geq 0}$ in $S_2$ only with the adjusted initial condition $t_1 = \frac{1}{2} - \delta$. The resulting solution then reads as

$$t_n = \frac{1}{2} - 2\delta + \delta\left(\frac{1}{2}\right)^{n-1}.$$

It is clear that $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$ converge and have their limits at $\frac{1}{2} + 2\delta$ and $\frac{1}{2} - 2\delta$ respectively. Consequently, the function $f$ is defined as

$$f(v) = \begin{cases} s_{\pi(v)}, & \text{if } v \in S_1 \\ t_{\pi(v)}, & \text{if } v \in S_2 \\ \frac{1}{2}, & \text{if } v = s \end{cases}, v \in V(G).$$

Clearly, $\delta$ may be chosen such that both $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$ are bounded from below by zero, and from above by 1. For such a $\delta$, it is easily checked that $|\,f\,| \in [0,1]$ and that the average condition is satisfied.

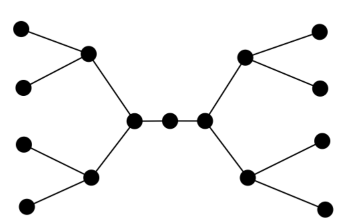

Figure 1: A partial embedding of the graph $G$, containing the root vertex and the resulting vertices of branching three times from the root.

# 5 Consequences of weaker assumptions in the Krein-Milman theorem

This section discusses the consequences for the statement of the Krein-Milman theorem when the assumptions in the theorem are weakened. Except for the local convexity requirement, it is clear that the statement will be false whenever the assumptions are weakened, as demonstrated by counterexamples in this section. For local convexity, the theorem will also fail as proved by Roberts in [9]. Firstly, let us consider weakening the assumption on compactness.

## 5.1 Weakening compactness: the binary intersection property and linearly closed sets

Trivially, for a non-empty convex subset $A$ of a locally convex Hausdorff space, the statement of the Krein-Milman will be false in general. A counterexample is provided by the space of real numbers $\mathbb{R}$ with respect to the standard topology, since $\mathbb{R}$ does not contain extreme points. The first consequence of weakening compactness is losing certainty of the set $A$ being closed, since compact sets in Hausdorff spaces are closed as shown in Lemma 2.22. Additionally, in some sense, boundedness is lost as well.

A natural question to ask is whether there exist locally convex Hausdorff spaces, in which some class of sets that are closed, bounded and convex, satisfy the Krein-Milman theorem, but are not compact. It turns out that such a class exists. Before considering such a class of sets, some necessary terminology has to be introduced. Let $A \subset E$ be closed, bounded and convex and define the *translations set* of $A$ as

$$\tau(A) = \{\, \lambda A + x \;:\; x \in E,\, \lambda \geq 0 \,\}.$$

Proceeding, a family of sets $\mathcal{A}$ is called *pairwise-intersecting* if $A_1 \cap A_2 \neq \varnothing$ for all $A_1, A_2 \in \mathcal{A}$. Lastly, a set $A$ is said to satisfy the *positive binary intersection property* if each pairwise-intersecting subfamily $\mathcal{A} \subset \tau(A)$ has a non-empty intersection. Nachbin proved in [8] that a closed, bounded and convex subset $A$ of a locally convex Hausdorff space $E$ satisfying the positive binary intersection theorem, also satisfies the statement of the Krein-Milman theorem.

Another class of non-compact sets satisfying (a slight variation of) the Krein-Milman theorem, has been studied by Klee in [7]. In his study, Klee generalized the statement of the Krein-Milman theorem to sets that are locally compact, closed and convex. Klee introduces the concept of a *linearly closed* set, which he uses in the proof of the two main results. Before defining what is meant with linearly closed sets, recall the definition of a *line*. A line from $a$ through $b$ (with $a, b \in E$) is the set $\{\, a + \lambda b \;:\; \lambda \in \mathbb{R} \,\}$. A convex set $A$ is called linearly closed if its intersection with every line is closed.

Additionally, Klee introduced the concept of an *extremal ray*. Recall that a *ray* or *half-line* emitting from the *source* $a$ through the point $b$ (with $a, b \in E$) is the set $\{ a + \lambda b : \lambda > 0 \}$. Now, an extremal ray of the set $A$, is a ray $\rho \subset A$ such that the following implication is valid: if the line segment $(a, b) \subset A$ intersects $\rho$, then $(a, b) \subset \rho$. Observe that if $\rho$ is an extremal ray of $A$ emitting from $a \in A$, then $a$ is an extreme point of $A$. The set of extremal rays is denoted by rext $A$.

The version of the Krein-Milman theorem satisfied by the class of linearly closed convex sets is marginally different from the formal statement. Klee proves two results, the first assuming that $E$ is finite dimensional and the second assuming that the set $A$ is locally compact. In the statements, another constraint is enforced on the set $A$, namely that it contains no lines. It is demonstrated by Klee that if $A$ contains a line, then it cannot have extreme points. The proofs of both statements can be found in [7].

**Theorem 5.1** (Klee - I)**.** *If $E$ is finite dimensional and $A \subset E$ a linearly closed and convex set containing no line, then* $A = \text{conv}\,(\text{ext}\, A \cup \text{rext}\, A)$.

Observe that in this statement the space $E$ is not required to be Hausdorff or locally convex. Proceeding, the second (and main) result proved by Klee is the next statement.

**Theorem 5.2** (Klee - II)**.** *Let $E$ be locally convex and Hausdorff. If $A$ is a locally compact closed convex set, containing no line, then* $A = \overline{\text{conv}}\,(\text{ext}\, A \cup \text{rext}\, A)$.

Let us give an example in which Theorem 5.1 can be applied, but the Krein-Milman theorem can not. Equip $\mathbb{R}^2$ with the standard topology and consider the closed first quadrant, which is the set

$$Q_1 = \{ (x, y) \in \mathbb{R}^2 \ : x \geq 0 \text{ and } y \geq 0\}.$$

In this context the Krein-Milman theorem would fail, since $Q_1$ is clearly not compact. However, $Q_1$ is convex, linearly closed and contains no line. It contains only one extreme point, which is the origin. The extremal rays are exactly the positive parts of the vertical and horizontal axes. Hence, by Theorem 5.1, the closed first quadrant is the convex hull of union of these and the origin. In other words, any point in the first quadrant may be written as a convex combination of points of the form $(0, a)$ and $(a, 0)$, for $a \in \mathbb{R}$. This is, of course, a trivial result. However, it demonstrates the benefit of Theorem 5.1.

## 5.2 Existence of extreme points in non-convex sets

The Krein-Milman theorem may be split into two independent results. Namely, the statement that a compact convex subset $A$ of a locally convex Hausdorff space $E$ contains extreme points, and the statement that $A$ is equal to the closed convex hull of its extreme points, where the first result is a consequence of the second. Clearly, the second assertion fails whenever $A$ is not convex. An easy counterexample would be given by a non-convex polygon $P \subset \mathbb{R}^2$, with respect to the standard topology. Trivially, the extreme points of a polygon form a subset of its vertices. Then trivially, the convex hull of the vertices would be a convex polygon, which is not equal to $P$. However, the first consequence of the Krein-Milman theorem is still valid.

Indeed, compact subsets of locally convex Hausdorff spaces contain extreme points. Observe that the proof of existence of extreme points in the Krein-Milman theorem (Theorem 3.2) does not use that the set $A$ is convex. It is important to notice that extreme points of compact sets are not as interesting as extreme points of convex sets. Due to the lack of structural theorems, as the Krein-Milman theorem, for non-convex sets, the study of extreme points of such sets is less interesting. However, acquiring information of extreme points of compact sets is not entirely worthless. For instance, continuous linear functionals attain their supremum or infimum in extreme points of compact sets. This can be deduced from Lemma 3.1.

## 5.3 Counterexample for non-locally convex spaces

The least obvious requirement needed for the Krein-Milman theorem is the local convexity of the space $E$. In the proof presented in this thesis, the Hahn-Banach separation theorem would not be applicable if the space were not locally convex. Therefore the presented proof would fail in this context. The counterexample (and its proof) presented by Roberts in [9] is rather complicated. Therefore, we will only construct the compact convex set in a non-locally convex space that does not contain extreme points, and not prove that it is a sufficient counterexample.

Firstly, some necessary terminology and notation will be introduced. Let $\mathcal{F}$ be the set of all functions $f : [0, 1] \to \mathbb{R}$ that are finite linear combinations of indicator functions functions of the form $\mathbb{1}_{[a,b)}$, where $a, b \in [0, 1]$. Observe that the functions in $\mathcal{F}$ are essentially simple functions, with an extra restriction on the sets on which the indicator functions are non-zero. Then $\mathcal{F}$ is a vector space with respect to the usual addition and scalar multiplication. Equipping the interval $[0, 1)$ with the Lebesgue measure $\lambda$ (and its corresponding $\sigma$-algebra ), define the norm $|| \cdot ||$ on $\mathcal{F}$ as $|| f || = \int |f| d\lambda$. Observe that the integration is valid since the functions in $\mathcal{F}$ are measurable. Proceeding, let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of partitions of the interval $[0, 1]$ such that each $I \in \pi_n$ is of the form $[a, b)$ (with $a, b \in [0, 1]$), all $I \in \pi_n$ have equal measure and the partition $\pi_n$ refines $\pi_{n+1}$. Additionally, define the sequence of vector subspaces $(V_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$ as $V_n = \text{span}\{ \mathbb{1}_A : A \in \pi_n \}$. Lastly, let $D = \{ f \in \mathcal{F} : f \geq 0, \text{ and } \int f d\lambda \leq 1 \}$ and define the sequence of sets $(E_n)_{n\in\mathbb{N}} = V_n \cap D$.

Let $V$ be the completion of the vector subspace $\bigcup_{n=1}^{\infty} V_n$ of $\mathcal{F}$. Roberts shows that $V$ is a non-locally convex Hausdorff space. In this space, the set $\overline{E}$, which is the closure of $E = \bigcup_{n=1}^{\infty} E_n$, is convex and compact in $V$, but does not contain extreme points.

## References


[1] Hermann Minkowski. *Gesammelte Abhandlungen*. B.G. Teubner, 1911.

[2] Mark Krein and David Milman. On extreme points of regular convex sets. *Studia Mathematica*, 9:133–138, 1940.

[3] Louis De Branges. The Stone-Weierstrass Theorem. *Proceedings of the American Mathematical Society*, 10(5):822–824, 1959.

[4] Vladimir Kadets. *A Course in Functional Analysis and Measure Theory*. Springer, 2018.

[5] Robert Ralph Phelps. *Lectures on Choquet's Theorem*. Springer, 2001.

[6] J. G. Mauldon. Extreme Points of Convex Sets of Doubly Stochastic Matrices. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 13:333–337, 1969.

[7] V. L. Klee Jr. Extremal structure of convex sets. *Archiv der Mathematik*, 8:234–240, 1957.

[8] Leopoldo Nachbin. Sur l'abondance des points extremaux d'un ensemble convexe borne et ferme. *Anais da Academia Brasileira de Ciências*, 34(3):445–448, 1962.

[9] James Roberts. A compact convex set with no extreme points. *Studia Mathematica*, 60(3):255–266, 1977.

[10] Vladimir Bogachev and Oleg Smolyanov. *Topological Vector Spaces and Their Applications*. Springer, 2017.

[11] Lawrence Narici and Edward Beckenstein. *Topological Vector Spaces*. CRC Press LLC, 2010.

[12] Rajnikant Sinha. *Real and Complex Analysis*. Springer, 2018.

[13] Marshall Harvey Stone. The Generalized Weierstrass Approximation Theorem. *Mathematics Magazine*, 21(4):167–184, 1948.

[14] F. Treves. *Topological Vector Spaces, Distributions and Kernels*. Elsevier Science, 1967.

[15] Serge Bernstein. Sur les fonctions absolument monotones. *Acta Mathematica*, 52:1–66, 1929.

[16] David G. Kendall. On Infinite Doubly-Stochastic Matrices and Birkhoffs Problem 111. *Journal of the London Mathematical Society*, 35:81–84, 1960.