Department of Mathematical Sciences
UNIVERSITY OF COPENHAGEN

June 5, 2020

Choquet Theory

Representation of compact convex sets

Christian Jonas Peifer
Bachelor's Thesis in Mathematics

Advisor: Magdalena Elena Musat

Abstract

The thesis presents an exposition of the main existence and uniqueness theorems of Choquet theory. First we introduce the concept of representing measures and establish their fundamental properties. We then prove Choquet's existence theorem for representing measures supported by the extreme points in the metrizable case. This is followed by a proof of the more general Choquet-Bishop-de Leeuw existence theorem, which abandons the assumption of metrizability for a slightly weaker conclusion. The theorem requires a further developing of theoretical tools involving the notion of maximal measures.

After a brief introduction to vector lattices and the decomposition lemma, we introduce the important idea of a Choquet simplex. We then prove some of its basic properties and see some important examples. This, along with a revisiting of maximal measures and the upper envelope, leads to a proof of the Choquet-Meyer uniqueness theorem for representing measures. We then discuss different notions of simplices and conclude the thesis by applying the established theorems to ergodic theory.

Contents

Introduction

This bachelor's thesis presents a survey and exposition of Choquet theory, which is the study of how compact convex sets can be represented by its set of so called extreme points. The theory rose out of potential theory and was intitally developed by Gustave Choquet in the 1950's. It has since then been heavily expanded upon by many others, both in its own theoretical depth and in its applications to many other areas of analysis.

The thesis is divided into seven sections. After establishing the needed preliminaries in section 1, the language of representing measures along with some properties of the important *resultant map* is developed in section 2. Section 3 introduces the concept of the upper envelope and proves Choquet's existence theorem for the metrizable case. In section 4 we introduce a certain ordering on measures, allowing for a proof of the more general Choquet-Bishop-de Leeuw existence theorem. In section 5 the notion of a Choquet simplex is developed, which is the key object in the Choquet-Meyer uniqueness theorem for representing measures of section 6. The last section then explains how all those theorems find applications in ergodic theory.

The theory is developed very thoroughly and in great detail. Assuming only an acquaintance with classical theorems and concepts from functional analysis, the thesis takes a bottom-up approach and thus strives to be as self contained as possible while still maintaining a clear direction through the material. It is therefore well-suited for an undergraduate student in the last year of study or a more experienced reader seeking an easy introduction to the subject.

Choquet theory has many of the attributes usually ascribed to a good mathematical theory. It is fundamentally about structure; it recognizes and studies how objects may be reconstructed from fewer and more easily understood building blocks. It also showcases how abstraction can aid in seeing the essential more clearly: By studying sets only by the property of being compact and convex, one realizes how phenomena in many different areas of analysis can be seen as instantiations of the same reconstruction process described by Choquet theory.

Of all the resources listed in the references, the author is especially indebted to the monograph *Lectures on Choquet's theorem* by Robert R. Phelps, originally published in 1966. It is difficult to express enough appreciation for the clarity of presentation and wealth of content Phelps has managed to contain in a very small book.

This bachelor's thesis was written in the spring of 2020, a time in which the world was heavily occupied by the breakout of the corona virus. The mathematical beauty of the subject proved sufficient to distract the author from these chaotic times, and hopefully its beauty is presented clearly enough to distract the reader as well.

Acknowledgements

I want to express my deepfelt gratitude to my family for always being eager to provide support and care and to my girlfriend Cecilie for being very patient with me. I also want to enthusiastically thank my advisor Magdalena for guidance through this project and many 'hyggelige' meetings, first in person, and then on Zoom. The thesis would not have been possible without them.

1 Preliminaries

In this section we gather useful definitions and results which are used frequently in the rest of the thesis. We start with the usual definitions regarding convexity:

Definition 1.1. Let V be a vector space and let C be a subset of V. A function $f : C \to \mathbb{R}$ is *convex* if for all $x_1, x_2 \in C$ and $\alpha \in (0, 1)$ we have

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2).$$

If the inequality is reversed, f is *concave*, and if equality holds, f is *affine*. Furthermore, if the inequality is strict whenever $x_1 \neq x_2$, f is *strictly convex*, and strict concavity is defined analogously.

Remark. We immediately see that f is convex if and only if $-f$ is concave and that f is affine if and only if it is both convex and concave.

The central subject is integral representation of compact convex sets, and these sit naturally inside topological vector spaces. To be able to develop the theory, it is of paramount importance that we have sufficiently many linear functionals on our space to separate points and sets. The general setting in which this is always the case is that of a Hausdorff *locally convex topological vector space*, abbreviated as LCS.

Definition 1.2. A topological vector space E is *locally convex* if it exhibits a convex neighbourhood base at 0.

The definition is quite short, but these spaces have many deep properties. It is worth noting that the neighbourhood base at 0 completely determines the topology; as translations are homeomorphishms, translates of the neighbourhood base at zero form neighbourhood bases at every other point of E. One can show that a topological vector space is locally convex if and only if the topology is generated by a family of seminorms.

This leads us to a brief discussion of weak and weak* topologies on Banach spaces:

Definition 1.3. Let Y be a Banach space and let Y^* denote the set of norm-continuous linear functionals. The *weak topology* on Y is the initial topology with respect to the maps $y \mapsto |f(y)|$, one for each $f \in Y^*$. Similiary, the *weak* topology* on Y^* is the initial topology with respect to the maps $f \mapsto |f(y)|$, one for each $y \in Y$.

Remark. A Banach space in the weak or weak* toplogy is a prime example of a locally convex topological vector space. The toplogies are additionally Hausdorff, and it follows immediately from the definitions that $y \mapsto f(y)$ is continuous in the weak topology while $f \mapsto f(y)$ is continuous in the weak* topology. We usually denote these evaluation maps $\hat{f}$ and $\hat{y}$, respectively.

A slightly more subtle fact is that a net $\{y_\alpha\}_{\alpha \in A} \subseteq Y$ converges to y in the weak topology if and only if $f(y_\alpha) \to f(y)$ for every $f \in Y^*$ while a net $\{f_\alpha\}_{\alpha \in A} \subseteq Y^*$ converges to f in the weak* topology if and only if $f_\alpha(y) \to f(y)$ for every $y \in Y$. Especially the latter will be used very frequently, and we refer to it as the pointwise convergence characterisation of weak* convergence. For an exposition of weak topologies we refer the reader to chapter V in [2].

For our purposes we will for the most part not be interested in the technical properties of these spaces, but mainly invoke the following two quintessential theorems:

Theorem 1.4 (Hahn-Banach Extension Theorem)**.** *Let V be a vector space over $\mathbb{R}$ and let $p : V \to \mathbb{R}$ be a sublinear functional. Let M be a subspace of V and let $f : M \to \mathbb{R}$ be a linear functional which satisfies $f(x) \leq p(x)$ for all $x \in M$. Then we may extend f to a linear functional $F : V \to \mathbb{R}$ which restricts to f and is dominated by p on V.*

The Hahn-Banach extension theorem is fundamental for our endeavour and we shall use it quite frequently. The theorem itself does not use convexity, in fact it does not even mention a topology. But another theorem we apply even more frequently is the Hahn-Banach separation theorem.

Theorem 1.5 (Hahn-Banach Separation Theorem)**.** *Let E be a real Hausdorff LCS and let $A, B \subseteq E$ be disjoint, convex sets. If A is closed and B is compact, there exist a continuous linear functional $T : E \to \mathbb{R}$ and $r \in \mathbb{R}$ such that*

$$\sup T(A) < r < \inf T(B)$$

As the set $H = \{x \in E \mid f(x) = r\}$ is a closed hyperplane in E, one can geometrically interpret this theorem as establishing the existence of a closed hyperplane which strictly separates the two sets.

There is, however, one technical lemma regarding the interaction of compactness and local convexity, which will be useful from time to time:

Lemma 1.6. *Let X be a compact convex subset of a real Hausdorff LCS E and let $U \subseteq X$ be open in X. Then every $y \in U$ has a closed convex neighbourhood C such that $C \subseteq U$*

Proof. It is a general fact of locally compact Hausdorff spaces that we may find an open set V such that

$$y \in V \subseteq \overline{V} \subseteq U.$$

As V is open, $V = X \cap U$ for some open $U \subseteq E$. By local convexity we may find a neighbourhood $U' \subseteq U$ of y which is open and convex. Then

$$V' := X \cap U' \subseteq V$$

is open and convex, so $C := \overline{V'} \subseteq U$ has all the desired properties. □

Notation 1.7. For notational ease E will from now on *always* denote a real Hausdorff locally convex topological vector space and $X \subseteq E$ will *always* denote a compact convex subset of E. We furthermore always assume that X is non-empty. If we repeat the assumptions on X, it is meant as a reminder to the reader and a service in case the thesis is used as a reference.

For a set $A \subseteq E$ we let $\mathrm{co}(A)$ denote the convex hull and $\overline{\mathrm{co}}(A)$ denote the closed convex hull of A. Since the vector space operations are continuous, the closure of a convex set is once again convex. Therefore $\overline{\mathrm{co}}(A)$ may freely be interpreted as either the smallest closed convex set containing A or the closure of the convex hull of A.

We will also introduce the following notation for two important subsets of $C(X)$:

$$\mathcal{A} = \{f \in C(X) \mid f \text{ is affine}\}, \quad \mathcal{C} = \{f \in C(X) \mid f \text{ is convex}\}.$$

We note that we always consider $C(X)$ equipped with the supremum norm. We also remark that we adopt the convention from [3] that a neighbourhood of x contains x as an interior point, but need not be open itself.

We are now ready to move on to the measure theoretic preliminaries. We will exclusively be working with Radon measures, which we define in the following:

Definition 1.8. A Borel measure μ on a locally compact Hausdorff space Y is *Radon* if it satisfies

1. $\mu(K) < \infty$ for all compact $K \subseteq Y$
2. μ is outer regular, so for each $B \in \mathbb{B}(Y)$

$$\mu(B) = \inf\{\mu(U) \mid U \text{ open}, B \subseteq U\}$$

3. μ is inner regular on open sets, so for open U

$$\mu(U) = \sup\{\mu(K) \mid K \subseteq U, K \text{ compact}\}$$

We denote the set of Radon measures on Y by $\mathcal{M}(Y)$.

We will occasionally be using some of the many useful properties of Radon measures. We collect a few of them in the following proposition:

Proposition 1.9. *Let Y be a locally compact Hausdorff space, let $\mathbb{B}(Y)$ be the Borel σ-algebra on Y and let μ be a σ-finite Radon measure on Y. Let $A \in \mathbb{B}(Y)$ and $f \in L^1(\mu)$ with $f \geq 0$. Then*

1. *μ is inner regular on A.*
2. *The measure $\mu_A(B) := \mu(A \cap B)$ for $B \in \mathbb{B}(Y)$ is Radon.*
3. *The measure $(f \cdot \mu)(B) := \int_B f \mathrm{d}\mu$ for $B \in \mathbb{B}(Y)$ is Radon.*

We refer the reader to Proposition 7.5 and Exercises 7. and 8. on p. 220 in [3]. Since we include regularity in the term probability measure, we define it explicitly:

Definition 1.10. A *probability measure* on a compact space K is a Radon measure μ which satisfies $\mu(K) = 1$. We denote the set of probability measures on K by $\mathcal{M}^1(K)$ and note that $\mathcal{M}^1(K)$ is naturally embedded in $C(K)^*$ by identifying μ with the map $f \mapsto \int_K f \mathrm{d}\mu$.

We note that probability measures are in particular σ-finite Radon measures, so Proposition 1.9 applies.

We conclude this section by stating the Riesz representation theorem, which in conjunction with the Hahn-Banach extension theorem will be our main tool for producing measures.

Theorem 1.11 (Riesz representation theorem)**.** *Let K be a compact Hausdorff topological space and let $I : C(K) \to \mathbb{R}$ be a positive linear functional. Then there exists a unique Radon measure μ such that for each $f \in C(K)$,*

$$I(f) = \int_K f \mathrm{d}\mu.$$

The Hahn-Banach theorems and the Riesz representation theorem will be the bread and butter for achieving all results of this thesis. The proofs of these extraordinary theorems are beautiful and we refer the reader to [3] for the extension and representation theorems, while a proof of the separation theorem can be found on p. 130 in [10].

2 Representing measures and their properties

In this section we introduce some of the concepts and language that will be studied throughout the rest of the thesis. Many of the results can be found in chapter 1 of Phelps' monograph [8]. In the following E will always be a *real* LCS, and X is always a compact convex subset of E. Furthermore we let E^* denote the set of real continuous linear functionals on E. We begin with defining extreme points, which is the central concept of the thesis:

Definition 2.1. Let V be a vector space and let C be a convex subset of E. A point $x \in C$ is an *extreme point* of C if $x = \alpha y + (1-\alpha)z$ with $y, z \in C$ and $\alpha \in (0,1)$ implies $x = y = z$.

When verifying that x is an extreme point, it actually suffices to check the case when $\alpha = 1/2$. Indeed, if $\alpha \neq 1/2$, we may assume $\alpha < 1/2$, and by defining $y' = z + 2\alpha(y - z)$, it is easy to verify that $y' \in C$, $y' \neq x$ and $x = 1/2(y' + z)$. Thus the existence of a proper convex combination representing x is equivalent to x being the midpoint of two distinct points. Although we will not use this frequently, it can be a great convenience.

All subsequent developments can broadly be seen as trying to answer the question: *In which way can a compact convex set X be reconstructed from its extreme points?* It is often helpful to visualize compact convex sets in $\mathbb{R}^2$ or $\mathbb{R}^3$ when establishing theorems which seek to answer this question. The truly interesting applications of Choquet theory, however, commonly arise in inifinte-dimensional vector spaces. And many of these applications are in some way connected to the following important example:

Example 2.2. Let K be a compact Hausdorff topological space and let $C(K)^*$ be equipped with the weak* topology. Then set of probability measures on K, $\mathcal{M}^1(K)$, is a compact convex subset of $C(K)^*$ by the identification $\mu \mapsto I_\mu$ where $I_\mu(f) = \int_K f \mathrm{d}\mu$ for each $f \in C(K)$.

Proof. It is well known that the Banach space $C(K)^*$ taken in its weak* topology is locally convex and Hausdorff. By the Riesz representation theorem there is a bijective correspondence between positive linear functionals I with $I(1) = 1$ and $\mathcal{M}^1(K)$. Abusing notation, we write

$$\mathcal{M}^1(K) = \{I \in C(K)^* \mid I(1) = 1,\ I \text{ positive}\}.$$

Since evaluation at the constant function 1, $\hat{1}$, is weak* continuous, the set $\{I \in C(K)^* \mid I(1) = 1\} = \hat{1}^{-1}(\{1\})$ is closed. Furthermore, by the characterisation of weak* convergence as pointwise convergence, the property of being positive is preserved by weak* limits. Therefore $\mathcal{M}^1(K)$ is the intersection of two closed sets and hence closed itself. Since

$$|I_\mu(f)| = \left| \int_K f \mathrm{d}\mu \right| \leq \int_K |f| \mathrm{d}\mu \leq \|f\|_\infty \mu(K) = \|f\|_\infty ,$$

we see that $\|I_\mu\| \leq 1$. Thus $\mathcal{M}^1(K)$ is a closed subset of the unit ball of $C(K)^*$, which is compact in the weak* topology by the classical theorem of Banach-Alaoglu, a proof of which can be found on p. 115 in [10]. We conclude that $\mathcal{M}^1(K)$ is itself compact. Convexity of $\mathcal{M}^1(K)$ is immediate from the interpretation of μ as the positive linear functional I_μ. $\square$

We will return to the above example several times during the thesis, and we shall see later that the extreme points of $\mathcal{M}^1(K)$ are exactly the Dirac measures. Next we define representing measures:

Definition 2.3. Let K be a compact subset of E. A probability measure μ on K is said to *represent* $x \in E$ if $\mu(f) := \int_K f \mathrm{d}\mu = f(x)$ for all $f \in E^*$.

We also often say x is the *resultant* of μ and occassionally that x is the *barycenter* of μ. The notation $\mu(f)$ may seem strange at first, but becomes natural with time and will be used ubiquitously in the sequel.

Notice that we abuse notation in this definition: We integrate f, which is defined on E, against μ, which is a measure on K. The most simple remedy for this technicality is to integrate the restriction $f_{|K}$, which is continuous on K so the integral is well-defined. A slightly more complicated, but also often very useful way to deal with it is the following: As K is a subspace of E, it can easily be shown that the Borel σ-algebra $\mathbb{B}(K)$ generated by the subspace topology equals the trace σ-algebra of the Borel σ-algebra on E. Put less confusingly, $\mathbb{B}(K) = \mathbb{B}(E) \cap K$. Thus we may naturally extend μ to be a probability measure on E by defining $\mu(A) = \mu(A \cap K)$ for all $A \in \mathbb{B}(E)$. We shall often let these considerations remain implicit.

We will use the notion of representing measure to discuss how a compact convex set can be reconstructed from its extreme points. Therefore we formalise what we mean when saying that a measure is supported by some Borel set A:

Definition 2.4. Let Y be a topological space. A probability measure μ on Y is *supported* by the Borel set A if $\mu(Y \setminus A) = 0$.

We frequently combine this definition with the method of extending the domain of a measure described earlier: When $K \subseteq K' \subseteq E$ with K, K' compact and μ a probability measure on K, we may unreservedly regard it as a probability measure on K' supported by K. The most common application of this is when K' is some compact set and $K = \overline{\mathrm{Ext}(K)}$.

2.1 Properties of the resultant map

We are now ready to prove some basic properties of representing measures, which will later on allow us the reformulate and generalise the classical theorem of Krein-Milman. We begin with the fact that under fairly weak assumptions probability measures always have a resultant.

Proposition 2.5. *Let K be a compact subset of E such that $X = \overline{co}(K)$ is compact. Then for each $\mu \in \mathcal{M}^1(K)$ there exists a unique $x \in X$ such that x is represented by μ.*

Proof. Let $\mu \in \mathcal{M}^1(K)$. For each $f \in E^*$ define $H_f = \{x \in E \mid f(x) = \mu(f)\}$. The existence of a resultant in X is then equivalent to the statement

$$\bigcap_{f \in E^*} H_f \cap X \neq \emptyset.$$

Since f is continuous and $H_f = f^{-1}(\{\mu(f)\})$, we immediately see that each H_f is closed, implying that $\{H_f \cap X\}_{f \in E^*}$ is a collection of closed subsets of X. By compactness of X it then suffices to show that this collection has the finite intersection property.

Therefore, let $f_1, \ldots, f_n \in E^*$. We wish to show that $\bigcap_{i=1}^n H_{f_i} \cap X \neq \emptyset$. To this end, define $T : E \to \mathbb{R}^n$ by $y \mapsto (f_1(y), \ldots, f_n(y))$. As each coordinate is linear and continuous, so is T. Since continuous maps preserve compactness and linear maps preserve convexity, $T(X)$ is both compact and convex. The set in question is then non-empty if and only if the point $p = (\mu(f_1), \ldots, \mu(f_n))$ is an element of $T(X)$.

To show this, assume $p \notin T(X)$. Since $T(X)$ is compact and convex, we may by the Hahn-Banach separation theorem find $h \in (\mathbb{R}^n)^*$ such that $h(p) > \sup h(T(X))$. By the Riesz representation theorem for continuous linear functionals on a Hilbert space (Theorem 5.25 in [3]), we find $a \in \mathbb{R}^n$ such that $h = \langle a, \cdot \rangle$.

To establish a contradiction, define $g = \sum_{i=1}^n a_i f_i$, where a_i are the coordinates of a. Then we do the following calculation:

$$h(p) = \langle a, p \rangle = \sum_{i=1}^n a_i p_i = \sum_{i=1}^n a_i \mu(f_i) = \int_K \sum_{i=1}^n a_i f_i \mathrm{d}\mu = \int_K g \mathrm{d}\mu.$$

Furthermore, for each $x \in X$ we have

$$h(Tx) = \langle a, Tx \rangle = \sum_{i=1}^n a_i f_i(x) = g(x).$$

Thus we may reformulate $h(p) > \sup h(T(X))$ as $\int_K g\mathrm{d}\mu > \sup g(X)$. This is however impossible, since $\int_K g\mathrm{d}\mu \leq \sup g(K)\mu(K) \leq \sup g(X)$ as $K \subseteq X$ and $\mu(K) = 1$. We conclude that $p \in T(X)$, $\{H_f \cap K\}_{f\in E^*}$ has the finite intersection property and μ has a resultant.

Compared to existence, the uniqueness of the resultant is easy: Assume both $x, y \in E$ are represented by μ. Then $f(x) = \int_K f\mathrm{d}\mu = f(y)$ for all $f \in E^*$. It is a consequence of the Hahn-Banach extension theorem that the continuous linear functionals separate points, so $x = y$. □

Now that we have established existence of the resultant of each $\mu \in \mathcal{M}^1(K)$, it is natural to define the *resultant map* $r : \mathcal{M}^1(K) \to \overline{\text{co}}(K)$ by mapping μ to its resultant. The proposition above can then be seen as ensuring that this map is well defined.

Having endowed $\mathcal{M}^1(K)$ with the weak* topology, it seems natural to ask whether the resultant map is continuous with respect to this topology. This is indeed the case:

Lemma 2.6. *Let $K \subseteq E$ be compact such that $\overline{co}(K)$ is also compact. The resultant map $r : \mathcal{M}^1(K) \to \overline{co}(K)$ is weak* continuous and affine.*

Proof. Let $\{\mu_\alpha\}_{\alpha\in A}$ be a net converging to the probability measure $\mu \in \mathcal{M}^1(K)$ in the weak* topology. Letting $r(\mu_\alpha) = x_\alpha$ and $r(\mu) = x$, we wish to show that $\lim_\alpha x_\alpha = x$. We do this by arguing that every subnet itself has a further subnet converging to x. Since $\mathcal{M}^1(K)$ is compact, every subnet will automatically have a converging subnet. As subnets of subnets are again subnets, it suffices to show that every *converging* subnet converges to x.

So let $\{x_\beta\}_{\beta\in B}$ be a converging subnet with $\lim_\beta x_\beta = y$. Since $\lim_\alpha \mu_\alpha = \mu$, the subnet $\{\mu_\beta\}_{\beta\in B}$ also converges to μ. We establish $x = y$ by showing $r(\mu) = y$ and appealing to the resultant map being well-defined. Taking some $f \in E^*$, $f_{|K} \in C(K)$. By the characterization of weak* convergence as pointwise convergence, $\mu_\beta(f_{|K})$ converges to $\mu(f_{|K})$. But this implies

$$f(y) = f(\lim_\beta x_\beta) = \lim_\beta f(x_\beta) = \lim_\beta \mu_\beta(f_{|K}) = \mu(f_{|K}) = \mu(f).$$

Thus $y = r(\mu) = x$, so r is in fact weak* continuous. That it is also affine follows immediately from the fact that being a resultant is defined in terms of $f \in E^*$, which are linear on E so in particular affine on K. □

The following proposition establishes a bridge between the concept of the closed convex hull and that of representing measures:

Proposition 2.7. *Let $K \subseteq E$ be compact. Then $\overline{co}(K) = r(\mathcal{M}^1(K))$. That is, x is in the closed convex hull of K if and only if x is represented by a probability measure on K.*

Proof. Assume first that x is represented by a probability measure μ on K. Assume by contradiction that $x \notin \overline{\text{co}}(K)$. But $\overline{\text{co}}(K)$ is closed, convex and disjoint from the compact convex singleton $\{x\}$, so by the Hahn-Banach separation theorem we may find $f \in E^*$ such that $f(x) > \sup f(\overline{\text{co}}(K))$. However, as x is represented by μ, we have

$$f(x) = \mu(f) = \int_K f\mathrm{d}\mu \leq \sup f(K)\mu(K) \leq \sup f(\overline{\text{co}}(K)).$$

Since this violates the choice of f, we conclude $x \in \overline{\text{co}}(K)$.

Conversely, assume $x \in \overline{\text{co}}(K)$. Then there exists a net $\{x_\alpha\}_{\alpha\in A} \subseteq \text{co}(K)$ which converges to x. We may write each x_α as $x_\alpha = \sum_{i=1}^{n_\alpha} a_i^\alpha y_i^\alpha$ with $\sum_{i=1}^{n_\alpha} a_i^\alpha = 1$ and $a_i^\alpha > 0$.

The following calculation then shows that x_α is represented by $\mu_\alpha := \sum_{i=1}^{n} a_i^\alpha \delta_{y_i^\alpha}$, where $\delta_{y_i^\alpha}$ is the Dirac measure of y_i^α. Letting $f \in E^*$,

$$\mu_\alpha(f) = \int_K f\mathrm{d}\mu_\alpha = \sum_{i=1}^{n} a_i^\alpha \int_K f\mathrm{d}\delta_{y_i^\alpha} = \sum_{i=1}^{n} a_i^\alpha f(y_i^\alpha) = f\left(\sum_{i=1}^{n} a_i^\alpha y_i^\alpha\right) = f(x_\alpha).$$

Hence we have a net $\{\mu_\alpha\}_{\alpha\in A} \subseteq \mathcal{M}^1(K)$, which is weak* compact by Example 2.2. This guarantees the existence of a converging subnet $\{\mu_\beta\}_{\beta\in B}$ whose limit we denote by μ. Just as we argued in the preceding lemma, μ represents x: If $f \in E^*$, then

$$\mu(f) = \lim_\beta \mu_\beta(f) = \lim_\beta f(x_\beta) = f(\lim x_\beta) = f(x).$$

Here we have used the fact that μ_β represents x_β and that f is continuous. □

Before we reformulate the Krein-Milman theorem, we translate one last concept from conventional convex analysis to the language of representing measures:

Proposition 2.8 (Bauer). *Let X be a compact convex set. Then $x \in X$ is an extreme point if and only if δ_x is the only probability measure on X representing x.*

Proof. Assume first that x is not an extreme point of X. Then $x = \alpha y + (1-\alpha)z$ for some $y, z \in X$ with $x \neq y \neq z$ and $\alpha \in (0,1)$. We note that δ_y and δ_z trivially represent y and z, respectively. Defining $\mu = \alpha\delta_y + (1-\alpha)\delta_z$ and using the fact from Lemma 2.2 that the resultant map r is affine, we immediately see that μ represents x. It is also obvious that $\mu \neq \delta_x$; the two measures differ on, e.g, $\{x\}$.

Assume next that x is an extreme point of X and that μ is a probability measure on X which represents it. We want to show that $\mu = \delta_x$. This is done by showing that μ is supported by $\{x\}$, since δ_x is the only probability measure on X satisfying this condition. To this end we use inner regularity of μ.

Let $K \subseteq X \setminus \{x\}$ be compact. To establish a contradiction, assume $\mu(K) > 0$. First we show that there exists $y \in K$ such that $\mu(B) > 0$ for every Borel neighbourhood B of y. If this were not the case, we could for each $y \in K$ find a Borel neighbourhood B_y such that $\mu(B_y) = 0$. Then $\{\mathrm{int}(B_y)\}_{y\in K}$ is an open cover of K, so we may find a finite subcover indexed by $y_1, \dots, y_n$, which yields

$$\mu(K) \leq \mu\left(\bigcup_{i=1}^n \mathrm{int}(B_{y_i})\right) \leq \mu\left(\bigcup_{i=1}^n B_{y_i}\right) \leq \sum_{i=1}^n \mu(B_{y_i}) = 0.$$

This however contradicts $\mu(K) > 0$, so we are guaranteed the existence of such y.

By Lemma 1.6 we may find a closed convex neighbourhood C of y such that $C \subseteq X \setminus \{x\}$. By our delicate choice of y we have $\mu(C) > 0$. Furthermore $\mu(C) < 1$, as we otherwise might regard μ as a probability measure on C which is compact and convex, so by Proposition 2.7 its resultant x is in C, contradicting $C \subseteq X \setminus \{x\}$.

Letting $\alpha := \mu(C) \in (0,1)$, we may define two new probability measures on X by $\mu_1(B) = \alpha^{-1}\mu(B \cap C)$ and $\mu_2(B) = (1-\alpha)^{-1}\mu(B \cap (X \setminus C))$ for each $B \in \mathbb{B}(X)$. These evidently measure to 1 and are Radon by Proposition 1.9, and by construction we have $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. Once again using that the resultant map is affine, we finally obtain

$$x = r(\mu) = r(\alpha\mu_1 + (1-\alpha)\mu_2) = \alpha r(\mu_1) + (1-\alpha)r(\mu_2).$$

Using Proposition 2.7 we see that $r(\mu_1) \in C \subseteq X \setminus \{x\}$, so x is expressed as a proper convex combination, contradicting the fact that it is extreme. Thus we may finally conclude that our assumption of $\mu(K) > 0$ is false, and by inner regularity we see that $\mu(X \setminus \{x\}) = 0$. Hence $\mu = \delta_x$ as desired. □

2.2 Reformulation of classical theorems

We are now ready to reformulate the Krein-Milman theorem in terms of representing measures. We recall that the classical Krein-Milman theorem states that when X is a non-empty compact convex set in a LCS, $X = \overline{\mathrm{co}}(\mathrm{Ext}(X))$. We refer the reader to p. 142 in [2] for a proof.

Theorem 2.9 (Reformulated Krein-Milman). *Let $X \subseteq E$ be compact and convex. Then each $x \in X$ is represented by a probability measure supported by $\overline{\mathrm{Ext}(X)}$.*

Proof. We show that this statement is equivalent to the usual formulation of the Krein-Milman theorem. Formally, the above theorem states $X \subseteq r(\mathcal{M}^1(\overline{\mathrm{Ext}(X)}))$, but Proposition 2.7 implies that equality holds. We also note that $\overline{\mathrm{co}}(\mathrm{Ext}(x)) = \overline{\mathrm{co}}(\overline{\mathrm{Ext}(X)})$, which follows easily from $\overline{\mathrm{co}}(\mathrm{Ext}(X))$ being a closed convex set containing $\overline{\mathrm{Ext}(X)}$. We then have

$$\overline{\mathrm{co}}(\mathrm{Ext}(X) = \overline{\mathrm{co}}(\overline{\mathrm{Ext}(X)} = r(\mathcal{M}^1(\overline{\mathrm{Ext}(X)})),$$

where we appeal to Proposition 2.7 for the second equality. Assuming the Krein-Milman theorem, the left hand side equals X and we see that $X = r(\mathcal{M}^1(\overline{\mathrm{Ext}(X)}))$, implying the statement of this theorem. Conversely, assuming this theorem we have $r(\mathcal{M}^1(\overline{\mathrm{Ext}(X)})) = X$, implying the classical Krein-Milman theorem. □

In the course of this thesis we will prove three theorems establishing the existence of representing measures which are in some sense supported by the extreme points. The reformulated Krein-Milman theorem is the first and coarsest of the three; it establishes the existence of a measure supported by the closure of the extreme points. In coarse topologies, such as the weak* topology on a dual space, there are few closed sets and hence a lot of information can be lost by taking closures. The following example originating from Example 9.5 in [11] illustrates this:

Example 2.10 (Unit ball in ℓ^2)**.** Let $X = \overline{B}(0,1) = \{x \in \ell^2(\mathbb{R}) \mid \|x\| \leq 1\}$ be the closed unit ball in the Hilbert space $\ell^2(\mathbb{R})$ equipped with the weak toplogy. X is convex by the triangle inequality, and since the weak and weak* topologies coincide on a Hilbert space, the Banach-Alaoglu theorem establishes compactness of X.

Any $x \in X$ with $\|x\| < 1$ can be expressed as a proper convex combination of two antipodal points, so we have $\mathrm{Ext}(X) \subseteq \{x \in \ell^2(\mathbb{R}) \mid \|x\| = 1\}$. To see that equality holds, let $x \in X$ with $\|x\| = 1$ and $x = 1/2(y+z)$ for $y, z \in X$ with $y \neq x$. By the triangle inequality we have $\|x\| \leq 1/2\,\|y\| + 1/2\,\|z\|$, which forces $\|y\| = \|z\| = 1$. An application of the parallelogram identity then yields

$$1 = \|x\|^2 = \|y+z\|^2/4 = (2(\|y\|^2 + \|z\|^2) - \|y-z\|^2)/4 = 1 - \|y-z\|^2.$$

From this we obtain $y = z$, i.e., $x \in \mathrm{Ext}(X)$.

We now have a compact convex set and have identified its extreme points, so we may apply the reformulated Krein-Milman theorem. Thus for each $x \in X$ there exists $\mu \in \mathcal{M}^1(X)$ which represents x and is supported by $\overline{\mathrm{Ext}(X)}$. This is, however, not very impressive: We have $\overline{\mathrm{Ext}(X)} = X$. Indeed, letting $x = (x_k)_{k\geq 1} \in X$, we may define a sequence $(x^{(n)})_{n\geq 1} \subseteq \ell^2(\mathbb{R})$ by

$$x^{(n)} = (x_1, x_2, \ldots, x_n, \sqrt{1 - \sum_{i=1}^n |x_i|^2}, 0, \ldots).$$

Then

$$\left\|x^{(n)}\right\|^2 = \sum_{i=1}^n |x_i|^2 + (1 - \sum_{i=1}^n |x_i|^2) = 1,$$

establishing $(x^{(n)})_{n\geq 1} \subseteq \mathrm{Ext}(X)$. Furthermore, by the Riesz representation theorem for continuous linear functionals on a Hilbert space and usual characterisation of weak convergence, we may verify $x^{(n)} \to x$ by showing $\langle x^{(n)}, y\rangle \to \langle x, y\rangle$ for every $y \in \ell^2(\mathbb{R})$. But for $y \in \ell^2(\mathbb{R})$ we have

$$|\langle x, y\rangle - \langle x^n, y\rangle| = |\langle x - x^{(n)}, y\rangle| = \left| x_{n+1} - \sqrt{1 - \sum_{i=1}^n |x_i|^2} \right| |y_{n+1}| + \sum_{i=n+2}^{\infty} |x_i y_i|,$$

which goes to 0 as $n \to \infty$. Hence we could have simply chosen $\mu = \delta_x$ as our representing measure supported by $\overline{\mathrm{Ext}(X)}$. We note that the argument above easily extends to $\ell^p(\mathbb{R})$ for $1 < p < \infty$ by using strict convexity of the map $\mathbb{R} \ni x \mapsto |x|^p$ instead of the parallelogram identity.

The tools we have constructed in this section give rise to a short and elegant proof of Milman's converse to the Krein-Milman theorem.

Theorem 2.11 (Milman)**.** *Let $X \subseteq E$ be compact and convex. If Z satisfies $X = \overline{co}(Z)$, then $Ext(X) \subseteq \overline{Z}$.*

Proof. Let $x \in \text{Ext}(X)$. By the assumption on Z and Propostion 2.7, x is represented by a measure supported by $\overline{Z}$. Then by Proposition 2.8 this measure is in fact δ_x, and δ_x being supported by $\overline{Z}$ implies $x \in \overline{Z}$. $\square$

We conclude this section by returning to Example 2.2 from the beginning of the section. It turns out that the Riesz representation theorem admits a nice interpretation in terms of representing measures. We first prove that the extreme points of $\mathcal{M}^1(K)$ are the Dirac measures, which is an interesting result on its own:

Proposition 2.12. *Let K be compact Hausdorff. Then $Ext(\mathcal{M}^1(K)) = \{\delta_x \mid x \in K\}$, and the embedding $x \mapsto \delta_x$ is a homeomorphism onto $Ext(\mathcal{M}^1(K))$. In particular, $Ext(\mathcal{M}^1(K))$ is closed.*

Proof. Assuming $\delta_x = \alpha\mu_1 + (1-\alpha)\mu_2$ for $\alpha \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{M}^1(K)$, we have $\mu_i(K \setminus \{x\}) = 0$ for $i \in \{1,2\}$. Since the μ_i's are additionally probability measures, this entails $\delta_x = \mu_1 = \mu_2$, from which we deduce $\delta_x \in \text{Ext}(\mathcal{M}^1(K))$.

To prove the reverse inclusion, we first note that if a probability measure μ only has one point such that all open neighbourhoods of this point have positive measure, then μ is the Dirac measure of the point. Indeed, denoting the point by x_0 and choosing open neighbourhoods U_x with $\mu(U_x) = 0$ for every $x \in K \setminus \{x_0\}$, let $C \subseteq \bigcup_{x \in X \setminus \{x_0\}} U_x$ be compact. Using compactness we obtain a finite subcover, which implies $\mu(C) = 0$. Inner regularity of μ then implies $\mu(K \setminus \{x_0\}) = 0$, and therefore $\mu = \delta_{x_0}$. A similar argument shows that if μ has zero such points, $\mu = 0$.

For the reverse inclusion suppose μ is not a Dirac measure. Then, by the above, there exist $x_1, x_2 \in K$ with $x_1 \neq x_2$ such that all neighbourhoods of the two points have positive measure. Separating x_1 and x_2 by open sets U and V, this implies $0 < \mu(U) < 1$, with the latter inequality being a consequence of $U^c \supseteq V$. We then define new probability measures on K by $\mu_1(B) = \mu(U)^{-1}\mu(B \cap U)$ and $\mu_2(B) = \mu(U^c)^{-1}\mu(B \cap U^c)$ for $B \in \mathbb{B}(K)$. This implies $\mu = \mu(U)\mu_1 + \mu(U)^c\mu_2$, showing that μ is not an extreme point of $\mathcal{M}^1(K)$.

The map $\varphi : K \to \text{Ext}(\mathcal{M}^1(K))$ defined by $\varphi(x) = \delta_x$ is surjective by the above and obviously injective. Since K is compact and $\text{Ext}(\mathcal{M}^1(K))$ is Hausdorff, it is sufficient to argue that φ is continuous to conclude that it is a homeomorphism. Letting $\{x_\alpha\}_{\alpha \in A}$ be a net converging to $x \in K$, we wish to prove $\{\delta_{x_\alpha}\}_{\alpha \in A}$ converges to δ_x. But by continuity of $f \in C(K)$,

$$\delta_x(f) = f(x) = \lim_\alpha f(x_\alpha) = \lim_\alpha \delta_{x_\alpha}(f),$$

which yields the desired result by the usual characterisation of weak* convergence. $\square$

For the following corollary we recall that a *discrete measure* is a measure of the form $\sum_{i=1}^n \alpha_i \delta_{x_i}$ for $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$. The discrete measures are exactly the convex hull of $\{\delta_x \mid x \in K\}$.

Corollary 2.13. *Let K be compact Hausdorff. Then the discrete measures are weak* dense in $\mathcal{M}^1(K)$. Equivalently, for each $\mu \in \mathcal{M}^1(K)$ there exists $m \in \mathcal{M}^1(\mathcal{M}^1(K))$ such that $m(\{\delta_x \mid x \in K\}) = 1$ and $m(\hat{f}) = \mu(f)$ for all $f \in C(K)$.*

Proof. The two statements are applications of the classical and reformulated Krein-Milman theorems to the compact convex set $X = \mathcal{M}^1(K)$. Since $\text{Ext}(\mathcal{M}^1(K)) = \{\delta_x \mid x \in K\}$, the first statement is a direct application.

For the second statement, we note that that the dual of $C(K)^*$ in the weak* topology consists precisely of the evaluation maps. We refer the reader to Theorem 1.3 on p. 125 in [2] for a short proof. Thus, for each $\mu \in \mathcal{M}^1(K)$, there exists $m \in \mathcal{M}^1(\mathcal{M}^1(K))$ which represents μ and is supported by $\{\delta_x \mid x \in K\}$. But m representing μ means that for each $f \in C(K)$ we have

$$m(\hat{f}) = \hat{f}(\mu) = \mu(f),$$

which is the desired conclusion. $\square$

Remark. It is interesting to note that the second part of the above corollary can be deduced without the reformulated Krein-Milman theorem. Indeed, let $\tilde{\mu}$ denote image measure of μ under the homeomorphism φ mapping $x \mapsto \delta_x$. Then we have $\tilde{\mu}(\hat{f}) = \mu(f)$ by the following exercise in evaluation maps:

$$\hat{f}(\varphi(x)) = \hat{f}(\delta_x) = \delta_x(f) = f(x).$$

So by identifying μ with $\tilde{\mu}$ we may pick m to be μ itself!

Even more interestingly, emphasizing the interpretation of μ as a positive linear function I_μ, any m supported by $\mathrm{Ext}(\mathcal{M}^1(K))$ having resultant I_μ descends to a measure on K representing I_μ in the sense of the Riesz representation theorem. Thus the uniqueness portion of Riesz' theorem essentially states that $m = \tilde{\mu}$ is the only choice in the above corollary. It is quite satisfying that by identification of μ with $\tilde{\mu}$, the two notions of representing I_μ coincide.

We see that $\mathcal{M}^1(K)$ has the interesting property that representing measures supported by the extreme points are unique. While the next two sections are devoted to studying the question of existence more thoroughly, we return to the question of uniqueness in sections 5 and 6. There we shall give precise conditions for unique representing measures, and we will see how this is intimately tied to the notion of a *Choquet simplex*, which $\mathcal{M}^1(K)$ is a prime example of. The idea of having measures on $\mathcal{M}^1(K)$ with certain properties is revisited in the last section, when we study applications of Choquet theory to invariant and ergodic measures.

3 Existence in the metrizable case

In this section we seek to prove Choquet's theorem, which is a sharpening of the reformulated Krein-Milman theorem in the case when X is metrizable. The proof of the main theorem and most of the auxiliary results originate from chapter 3 in [8]. The key idea in the proof is an application of the Hahn-Banach extension theorem to a delicately chosen sublinear functional. This functional is defined using the notion of an upper envelope, so we will introduce this concept and prove some of its basic properties.

3.1 The upper envelope

We start with two definitions:

Definition 3.1. Let Y be a topological space. A function $f : Y \to \mathbb{R}$ is upper semicontinuous if $f^{-1}((-\infty, r))$ is open for all $r \in \mathbb{R}$.

We note that any upper semicontinuous function is Borel measurable as the open rays generate $\mathbb{B}(\mathbb{R})$. It turns out that the notion of upper semicontinuity coincides with the hypograph $H_f = \{(x, r) \in E \times \mathbb{R} \mid f(x) \geq r\}$ being closed. We prove only one part of the equivalence, since this suffices for our purposes:

Lemma 3.2. *Let Y be a topological space and let $f : Y \to \mathbb{R}$ be upper semicontinuous. Then the hypograph of f is closed.*

Proof. Let $\{(x_\alpha, r_\alpha)\}_{\alpha \in A} \subseteq H_f$ with $\lim_\alpha (x_\alpha, r_\alpha) = (x, r)$. To see that $(x, r) \in H_f$ assume it is not, i.e, $f(x) < r$. We may then find $\varepsilon > 0$ such that $f(x) < r - \varepsilon$. Since f is upper semicontinuous, the set $f^{-1}((-\infty, r - \varepsilon))$ is open and contains x. By definition of the product topology we have $\lim_\alpha x_\alpha = x$ and $\lim_\alpha r_\alpha = r$, so $\{x_\alpha\}_{\alpha \in A}$ is eventually in $f^{-1}((-\infty, r - \varepsilon))$ and $\{r_\alpha\}_{\alpha \in A}$ is eventually in $(r - \varepsilon, \infty)$. Finding an α_0 such that both of these conditions are met, we see that $f(x_{\alpha_0}) < r - \varepsilon$ while $r_{\alpha_0} > r - \varepsilon$, contradicting that $(x_{\alpha_0}, r_{\alpha_0}) \in H_f$. $\square$

We are now ready to define the upper envelope itself. We recall that $\mathcal{A}$ denotes the set of continuous affine functions on K.

Definition 3.3. Let K be a compact subset of E and let $f : K \to \mathbb{R}$ be bounded. The *upper envelope* of f is defined by $\bar{f}(x) := \inf\{h(x) \mid h \in \mathcal{A},\ h \geq f\}$.

The upper envelope approximates a function locally by affine functions which are globally greater than the function. We will later see that for strictly convex functions, this approximation only attains equality on the extreme points of the domain. We want to provide the reader with this very clear image of the upper envelope from Figure 28.4 in [1]:

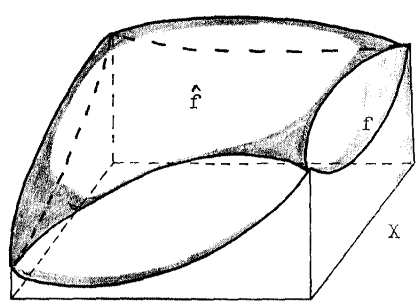

Figure 1: From p. 162 in *Lectures on Analysis* by G. Choquet

Before we prove some useful properties of the upper envelope, we establish this helpful lemma:

Lemma 3.4. *Let K be a compact subset of E, let $f : K \to \mathbb{R}$ be bounded and let $g : K \to \mathbb{R}$ be concave and upper semicontinuous with $g \geq f$. Then $g \geq \bar{f}$.*

Proof. Assume that the conclusion is false, i.e., there exists $x_0 \in K$ such that $g(x_0) < \bar{f}(x_0)$. As established in the previous lemma, g being upper semicontinuous implies that its hypograph H_g is closed, while concavity ensures that it is convex. The product space $E \times \mathbb{R}$ is also a real Hausdorff LCS, so we may apply the Hahn-Banach separation theorem to the closed convex set H_g and the compact set $\{(x_0, \bar{f}(x_0))\}$, which are disjoint by hypothesis. Hence we find $L \in (E \times \mathbb{R})^*$ and $\lambda \in \mathbb{R}$ such that

$$\sup L(H_g) < \lambda < L(x_0, \bar{f}(x_0)).$$

Note that in particular $L(x_0, g(x_0)) < L(x_0, \bar{f}(x_0))$, which implies that

$$0 < L(0, \bar{f}(x_0) - g(x_0)) = (\bar{f}(x_0) - g(x_0))L(0, 1).$$

We conclude that $L(0, 1) > 0$ and define $h : K \to \mathbb{R}$ by the equation $L(x, h(x)) = \lambda$. Rearranging terms one obtains the more direct expression

$$h(x) = \frac{\lambda - L(x, 0)}{L(0, 1)},$$

which immediately yields well-definedness and continuity of h. Since h is of the form $T + r$ for T linear and $r \in \mathbb{R}$, we also see that it is affine. For any $x \in K$ we have $(x, g(x)) \in H_g$, and furthermore

$$0 < \lambda - L(x, g(x)) = L(x, h(x)) - L(x, g(x)) = L(0, h(x) - g(x)) = (h(x) - g(x))L(0, 1).$$

We established earlier that $L(0, 1) > 0$, so this forces $h(x) > g(x)$. The same argument establishes that $h(x_0) < \bar{f}(x_0)$. Since $h \geq g \geq f$ we have found a continuous affine function dominating f with $\bar{f}(x_0) > h(x_0)$, yielding a contradiction. □

The next proposition contains a handful of extremely useful properties of the upper envelope which will be used very frequently in subsequent developments:

Proposition 3.5. *Let K be a compact Hausdorff space and let $f, g : K \to \mathbb{R}$ be bounded. Then the upper envelopes of f and g enjoy the following properties:*

1) $\bar{f}$ is concave, bounded and upper semicontinuous

2) If f is concave and upper semicontinuous, $f = \bar{f}$.

3) $\overline{f + g} \leq \bar{f} + \bar{g}$ and $|\bar{f} - \bar{g}| \leq \|f - g\|_\infty$

4) If furthermore $g \in \mathcal{A}$ and $r \geq 0$, then $\overline{f + g} = \bar{f} + g$ and $\overline{rf} = r\bar{f}$

Proof. The proposition states many properties, many of which follow in a straightforward manner from the definition of the upper envelope. We will adjust the level of detail in the proof accordingly.

Proof of 1) To prove $\bar{f}$ is concave let $x, y \in K$ and $\alpha \in (0, 1)$. We note that

$$\{\alpha h_1(x) + (1 - \alpha)h_2(y) \mid h_1, h_2 \in \mathcal{A}, h_1, h_2 \geq f\} \supseteq \{\alpha h(x) + (1 - \alpha)h(y) \mid h \in \mathcal{A}, h \geq f\}.$$

Since $\bar{f}(\alpha x + (1 - \alpha)y)$ is the infimum of the right hand side, we find

$$\begin{aligned} \bar{f}(\alpha x + (1 - \alpha)y) &\geq \inf\{\alpha h_1(x) + (1 - \alpha)h_2(y) \mid h_1, h_2 \in \mathcal{A}, h_1, h_2 \geq f\} \\ &= \alpha \inf\{h_1(x) \mid h_1 \geq f, h_1 \in \mathcal{A}\} + (1 - \alpha) \inf\{h_2(y) \mid h_2 \geq f, h_2 \in \mathcal{A}\} \\ &= \alpha\bar{f}(x) + (1 - \alpha)\bar{f}(y), \end{aligned}$$

proving that $\bar{f}$ is concave. As f is bounded, $f \leq C$ for some $C > 0$. The constant function C is affine and continuous and therefore $\bar{f} \leq C$. Since $f \leq \bar{f}$ by definition, $\bar{f}$ is also bounded from below.

To prove $\bar{f}$ is upper semicontinuous, let $r \in \mathbb{R}$ and $x_0 \in \bar{f}^{-1}((-\infty, r))$. By definition of $\bar{f}$ we can find $h \in \mathcal{A}$ with $h \geq f$ such that $\bar{f}(x_0) < h(x_0) < r$. Since h is continuous, $h^{-1}((-\infty, r))$ is open and $x_0 \in h^{-1}((-\infty, r))$. Then, for any $x \in h^{-1}((-\infty, r))$,

$$\bar{f}(x) \leq h(x) < r,$$

which implies $h^{-1}((-\infty, r)) \subseteq \bar{f}^{-1}((-\infty, r))$. Hence $\bar{f}$ is indeed upper semicontinuous.

Proof of 2) When f is concave and upper semicontinuous, Lemma 3.4 applies to f itself, so $f \geq \bar{f}$. Since we also have $f \leq \bar{f}$, this implies $f = \bar{f}$.

Proof of 3) $\overline{f+g} \leq \bar{f} + \bar{g}$ is similar to concavity: We have

$$\begin{aligned}(\bar{f} + \bar{g})(x) &= \inf\{h_1(x) \mid h_1 \geq f, h_1 \in \mathcal{A}\} + \inf\{h_2(x) \mid h_2 \geq g, h_2 \in \mathcal{A}\} \\ &= \inf\{(h_1 + h_2)(x) \mid h_1 \geq f, h_2 \geq g, h_1, h_2 \in \mathcal{A}\}.\end{aligned}$$

Since $h_1 \geq f$ and $h_2 \geq g$, $h_1 + h_2 \geq f + g$, which means that $\overline{f+g}$ is the infimum of a superset of the set defining $\bar{f} + \bar{g}$, hence $\overline{f+g} \leq \bar{f} + \bar{g}$. Using this, we see that

$$\bar{f} = \overline{f - g + g} \leq \overline{f-g} + \bar{g}.$$

Thus $\bar{f} - \bar{g} \leq \overline{f-g} \leq \|f-g\|_\infty$, where we use that the constant function $\|f-g\|_\infty$ is continuous and affine. Exchanging the roles of f and g yields

$$\bar{g} - \bar{f} = -(\bar{f} - \bar{g}) \leq \|g - f\|_\infty = \|f - g\|_\infty,$$

so $|\bar{f} - \bar{g}| \leq \|f-g\|_\infty$.

Proof of 4) Assume $g \in \mathcal{A}$. Then $g = \bar{g}$ by 2), and using 3) yields

$$\overline{f+g} \leq \bar{f} + \bar{g} = \bar{f} + g.$$

The other inequality follows from the fact that if $h \in \mathcal{A}$ and $h \geq f + g$, then $h - g \geq f$. This implies $\bar{f} \leq h - g$ for all such h, so

$$\bar{f}(x) \leq \inf\{h(x) - g(x) \mid h \in \mathcal{A}, h \geq f + g\} = \inf\{h(x) \mid h \in \mathcal{A}, h \geq f+g\} - g(x) = \overline{f+g}(x) - g(x).$$

The very last assertion might be the easiest: If $r = 0$ we have $\bar{0} = 0$ by 2). Otherwise let $r > 0$. Then for $h \in \mathcal{A}$ we have $h \geq rf$ if and only if $h/r \geq f$, so $\bar{f} = \overline{rf}/r$. □

3.2 Choquet's theorem (Existence)

The proof of Choquet's theorem is built around the existence of a strictly convex continuous function on X. The following lemma establishes that such a function always exists, given metrizability:

Lemma 3.6. *Let X be a compact convex metrizable subset of E. Then there exists a strictly convex continuous function $g : X \to \mathbb{R}$.*

Proof. Since X is compact Hausdorff, it is in particular completely regular which coupled with metrizability implies $C(X)$ being separable. The proof of this is slightly involved, and we refer the reader to Theorem 6.6 on p. 140 in [2]. Since any subset of a separable metric space is separable, the subset

$$\mathcal{A}^1 = \{h \in C(X) \mid h \in \mathcal{A}, \|h\|_\infty = 1\}$$

is separable. It also separates points, since if $x, y \in X$ with $x \neq y$, we can find $f \in E^*$ such that $f(x) \neq f(y)$. Then $f' := f_{|X} / \left\|f_{|X}\right\|_\infty \in \mathcal{A}^1$ and $f'(x) \neq f'(y)$.

Now let $\{h_n \mid n \in \mathbb{N}\}$ be a dense subset of $\mathcal{A}^1$. This will also separate points, since if $h_n(x) = h_n(y)$ for all $n \in \mathbb{N}$, then for any $f \in \mathcal{A}^1$ we can find $(h_{n_k})_{k\geq 1}$ with $\lim_{k\to\infty} h_{n_k} = f$. This implies

$$f(x) = \lim_{k\to\infty} h_{n_k}(x) = \lim_{k\to\infty} h_{n_k}(y) = f(y).$$

Since $\mathcal{A}^1$ separates points, we see that $\{h_n \mid n \in \mathbb{N}\}$ does as well.

Now define $g = \sum_{n=1}^{\infty} 2^{-n} h_n^2$. This uniform limit exists since $C(X)$ is complete, and

$$\left\| \sum_{n=N}^{M} 2^{-n} h_n^2 \right\|_\infty \leq \sum_{n=N}^{M} 2^{-n} \left\| h_n^2 \right\|_\infty = \sum_{n=N}^{M} 2^{-n}$$

when $N \leq M$, which implies the Cauchy criterion.

We furthermore claim that g is strictly convex. Let $x, y \in X$ with $x \neq y$ and $\alpha \in (0,1)$. Then $h_{n_0}(x) \neq h_{n_0}(y)$ for some $n_0 \in \mathbb{N}$. Since the map $t \mapsto t^2$ from $\mathbb{R} \to \mathbb{R}$ is strictly convex,

$$\alpha h_{n_0}^2(x) + (1-\alpha) h_{n_0}^2(y) > (\alpha h_{n_0}(x) + (1-\alpha) h_{n_0}(y))^2 = (h_{n_0}(\alpha x + (1-\alpha) y))^2 = h_{n_0}^2(\alpha x + (1-\alpha) y).$$

This inequality produces an $\varepsilon > 0$ such that

$$2^{-n_0} h_{n_0}^2(\alpha x + (1-\alpha) y) < 2^{-n_0}(\alpha h_{n_0}^2(x) + (1-\alpha) h_{n_0}^2(y)) - \varepsilon.$$

Since h_n^2 is convex for all other $n \in \mathbb{N}$, this implies that for $N > n_0$ we have

$$\begin{aligned}
\sum_{n=1}^{N} 2^{-n} h_n^2(\alpha x + (1-\alpha) y) &< \sum_{n=1}^{N} 2^{-n}(\alpha h_n^2(x) + (1-\alpha) h_n^2(y)) - \varepsilon \\
&= \alpha \sum_{n=1}^{N} 2^{-n} h_n^2(x) + (1-\alpha) \sum_{n=1}^{N} h_n^2(y) - \varepsilon.
\end{aligned}$$

Passing to the limit we see that g is strictly convex. $\square$

It is interesting to note that not only does metrizability entail the existence of a strictly convex continuous function, it is actually *equivalent* to the existence of such a function. Therefore the proof of Choquet's theorem does not readily generalise to the non-metrizable case. It will, however, still be useful in the general setting, as we shall see in the next section. For a proof of the equivalence we refer the reader to Theorem 10.56 in [6].

We now turn to proving that when X is metrizable, the set of extreme points is Baire and hence in particular Borel. Although there are several definitions of Baire sets in the literature, they coincide when X is compact Hausdorff, which is the only case that interests us. Recalling that G_δ sets are those that can be expressed as a countable intersection of open sets, we choose the following:

Definition 3.7. The Baire σ-algebra is the σ-algebra generated by the compact G_δ-sets.

Since G_δ-sets are in particular Borel, the Baire σ-algebra is of course contained in $\mathbb{B}(X)$. The following proof that $\mathrm{Ext}(X)$ is Baire is based on the proof of Theorem 10.7 in [11]:

Proposition 3.8. *Let X be compact, convex and metrizable. Then $\mathrm{Ext}(X)$ is Baire.*

Proof. Let $g : X \to \mathbb{R}$ be a strictly convex continuous function, the existence of which was established in the preceding lemma. For each $n, m \in \mathbb{N}$ define the set

$$G_{n,m} := \left\{ x \in X \mid \exists y, z \in X : x = \frac{1}{2}(y+z),\ g(x) < \frac{1}{2}(g(y)+g(z)) - \frac{1}{n} + \frac{1}{n+m} \right\}.$$

We argue that $G_{n,m}$ is open. The function $X \times X \to \mathbb{R}$ defined by

$$(y,z) \mapsto g((y+z)/2) - (g(y)+g(z))/2$$

is continuous as the composition of continuous functions. Letting A denote the preimage of $(-\infty, -1/n + 1/(n+m))$, A is then open in $X \times X$. Now let $x \in G_{n,m}$. Then there exists a pair $(y,z) \in A$ such that $x = 1/2(y+z)$. Since A is open, we may find open sets $U, V \subseteq X$ containing (y,z) such that $U \times V \subseteq A$. Then $1/2(y'+z') \in G_{n,m}$ for any $(y',z') \in U \times V$. Furthermore, since translations are homeomorphisms, $U+V=\bigcup_{u\in U} u+V$ is open, and hence $1/2(U+V)$ is open. Therefore $1/2(U+V)$ is an open set containing x which is itself contained in $G_{n,m}$, implying that x is an interior point. Thus each $G_{n,m}$ is open, and the sets

$$F_n := \bigcap_{m=1}^{\infty} G_{n,m} = \left\{ x \in X \mid \exists y,z \in X : x = \frac{1}{2}(y+z),\ g(x) \leq \frac{1}{2}(g(y)+g(z)) - \frac{1}{n} \right\}$$

are therefore G_δ. Now let B be the preimage of $(-\infty, -1/n]$ of the above defined map. Once again appealing to continuity, B is closed, and hence compact in $X \times X$. Then the image of B under the continuous map $(y,z) \mapsto 1/2(y+z)$ is compact and exactly equal to F_n. We conclude each F_n is compact G_δ and in particular Baire.

Now if x is *not* an extreme point, there exist $y,z \in X$ with $y \neq z$ such that $x = 1/2(y+z)$. Since g is strictly convex we have $g(x) < 1/2(g(y)+g(z))$, implying that x is in some F_n. Conversely, if $x \in F_n$ for some $n \in \mathbb{N}$, we have

$$g(x) < \frac{1}{2}(g(y)+g(z))$$

for some $y,z \in X$ with $x = 1/2(y+z)$. Then we cannot have $x = y = z$, so x is not an extreme. Hence

$$\mathrm{Ext}(X)^c = \bigcup_{n=1}^{\infty} F_n,$$

and we see that the extreme points are indeed Baire. □

We are now ready to prove Choquet's theorem:

Theorem 3.9 (Choquet)**.** *Let X be a compact, convex and metrizable subset of E. Then each $x \in X$ is represented by a measure supported by Ext(X).*

Proof. Let $x_0 \in X$ and let g be a strictly convex function on X, the existence of which is guaranteed by Lemma 3.6. We define a functional p on $C(X)$ by $p(f) = \bar{f}(x_0)$. Then p is subadditive since

$$p(f+f') = \overline{f+f'}(x_0) \leq \bar{f}(x_0) + \overline{f'}(x_0) = p(x) + p(y),$$

where we have used Proposition 3.5 for the inequality. Since positive homogeneity follows immediately from the same proposition, we infer that p is sublinear.

Furthermore we define a linear functional $T : \mathcal{A} + \mathbb{R}g \to \mathbb{R}$ by $T(h+rg) = h(x_0) + r\bar{g}(x_0)$. We now claim that p dominates T. First, let $r \geq 0$. Then

$$T(h+rg) = (h + r\bar{g})(x_0) = \overline{h+rg}(x_0) = p(h+rg),$$

where we use both statements in part 4) of Proposition 3.5. Since g is convex, rg will be concave when $r < 0$, so $\overline{rg} = rg$. This yields

$$T(h+rg) = h(x_0) + r\bar{g}(x_0) \leq h(x_0) + rg(x_0) = h(x_0) + \overline{rg}(x_0) = \overline{h+rg}(x_0) = p(h+rg).$$

For the inequality we used the fact that $g(x_0) \leq \bar{g}(x_0)$, which implies $rg(x_0) \geq r\bar{g}(x_0)$ when $r < 0$. We also used parts 2) and 4) of the aforementioned proposition.

The Hahn-Banach extension theorem allows us to extend T to a linear functional L on all of $C(X)$ which restricts to T on $\mathcal{A} + \mathbb{R}g$. To show that L is positive, we show that it is non-positive on non-positive functions. Let $f \leq 0$. Then, since $0 \in \mathcal{A}$, $\bar{f} \leq 0$. This implies

$$L(f) \leq p(f) = \bar{f}(x_0) \leq 0.$$

By the Riesz representation theorem there exists a Radon measure μ which satisfies $L(f) = \int_X f \mathrm{d}\mu$ for all $f \in C(X)$. We claim that this is indeed a probability measure representing x_0 which is supported by Ext(X). We see that

$$\mu(X) = \int_X 1 \mathrm{d}\mu = L(1) = T(1) = \bar{1}(x_0) = 1,$$

which means that μ is indeed a probability measure. Since the restriction of an $f \in E^*$ to X is in $\mathcal{A}$, we have $\int_X f \mathrm{d}\mu = T(x_0) = f(x_0)$. Thus μ does indeed represent x_0.

It remains to show that μ is supported by the extreme points of X. This is done by showing that it is supported by $\mathcal{E} = \{x \in X \mid g(x) = \bar{g}(x)\}$, i.e., the points of X where the strictly convex function g equals its upper envelope. We note that this set is clearly Borel as $\bar{g}$ and g are measurable. Let us first show that $\mathcal{E} \subseteq \mathrm{Ext}(X)$. Let $x \in \mathcal{E}$ and suppose $x = \alpha y + (1-\alpha)z$ for some $y \neq z$ and $\alpha \in (0,1)$. Then

$$g(x) = g(\alpha y + (1-\alpha)z) < \alpha g(y) + (1-\alpha)g(z) \leq \alpha\bar{g}(y) + (1-\alpha)\bar{g}(z) \leq \bar{g}(\alpha y + (1-\alpha)z) = \bar{g}(x),$$

where we use strict convexity of g, the fact that $g \leq \bar{g}$ and concavity of $\bar{g}$. This contradicts the fact that $x \in \mathcal{E}$, so $\mathcal{E} \subseteq \mathrm{Ext}(X)$.

To complete the proof we must show that μ is supported by $\mathcal{E}$. We do this by showing

$$\int_X (\bar{g} - g)\mathrm{d}\mu = 0,$$

since this implies that the non-negative function $\bar{g} - g$ is 0 almost everywhere, i.e., $\mu(\mathcal{E}^c) = 0$. We note that $\int_X \bar{g} - g\mathrm{d}\mu = 0$ is equivalent to $\mu(\bar{g}) = \mu(g)$, and see that $\mu(\bar{g}) \geq \mu(g)$ as $\bar{g} \geq g$. The other inequality takes some consideration. As

$$\mu(g) = T(g) = \bar{g}(x_0),$$

we can show $\mu(g) \geq \mu(\bar{g})$ by the definition of $\bar{g}(x_0)$. So let $h \in \mathcal{A}$ with $h \geq g$. By construction $\bar{g} \leq h$, so

$$\mu(\bar{g}) \leq \mu(h) = T(h) = h(x_0).$$

Therefore, since $\bar{g}(x_0)$ is the infimum over all such $h(x_0)$,

$$\mu(\bar{g}) \leq \bar{g}(x_0) = \mu(g).$$

This lets us conclude that $\mu(\mathcal{E}^c) = 0$. Finally, since the set of extreme points is Borel by Proposition 3.8, μ is supported by $\mathrm{Ext}(X)$. □

The upper envelope of a bounded function is always upper semicontinuous, and one may wonder if it is always continuous. We conclude this section with an example showing that this is not the case:

Example 3.10. Let $X = \overline{B}(0,1)$ be the closed unit ball in $\ell^2(\mathbb{R})$ just as in Example 2.10. The closed unit ball in a dual space is metrizable in the weak* topology if and only if its predual is separable, which is the case with $\ell^2(\mathbb{R})$. We refer the reader to Theorem 5.1 on p. 134 in [2] for a proof of this fact. Therefore X is compact, convex and metrizable, and Choquet's theorem applies to X and any of its closed convex subsets.

Now let $g : X \to \mathbb{R}$ be a strictly convex continuous function on X. The upper envelope $\overline{g}$ is then concave and upper semicontinuous by Proposition 3.5, and we know from Example 2.10 that $\overline{\mathrm{Ext}(X)} = X$. As we shall see much later in Proposition 6.7, g and $\overline{g}$ coincide on $\mathrm{Ext}(X)$. Now let $x \in X$ and let $(x_n)_{n\geq 1} \subseteq \mathrm{Ext}(X)$ be a sequence converging to x. Assuming $\overline{g}$ is continuous, we see that

$$\overline{g}(x) = \lim_{n\to\infty} \overline{g}(x_n) = \lim_{n\to\infty} g(x_n) = g(x).$$

Hence $g = \overline{g}$, and we see that g is both strictly convex and concave, which is clearly impossible.

4 Existence in the general setting

In this section we will work towards and prove the Choquet-Bishop-de Leeuw theorem, which states that every point of X has a representing measure which is supported by every Baire set containing the extreme points. While the conclusion is clearly weaker than Choquet's theorem, this theorem does not assume metrizability, extending its scope of applicability.

Furthermore, since we no longer have a strictly convex function on X, we will need a new set of ideas to solve this more general problem. Many of these revolve around an equivalence relation on $\mathcal{M}^1(X)$ and a partial order on $\mathcal{M}(X)$, and how these relate to the upper envelope. Before we introduce the equivalence relation, we remind the reader that X always denotes a non-empty compact convex set in a real Hausdorff LCS E, and that we let $\mathcal{A}$ and $\mathcal{C}$ denote the set of continuous affine and convex functions on X, respectively. We refer the reader to chapter 4 of [8] as the source of many of this section's results.

4.1 Equivalence of measures

Definition 4.1. Let $\mu, \nu \in \mathcal{M}^1(X)$. Then $\mu \sim \nu$ if $\mu(f) = \nu(f)$ for all $f \in \mathcal{A}$

We note that linear maps are in particular affine, so $\mu \sim \nu$ implies that μ and ν have the same resultant, i.e., $r(\mu) = r(\nu)$, where r is the resultant map. One commonly thinks of continuous affine functions as having the form $f + c$ for some (continuous) linear f and $c \in \mathbb{R}$. While there exist affine functions which cannot be characterised this way, we have the following density result:

Proposition 4.2. *The subspace* $E^*_{|X} + \mathbb{R} = \{f_{|X} + c \mid f \in E^*, c \in \mathbb{R}\}$ *is uniformly dense in* $\mathcal{A}$.

Proof. Since pointwise limits of affine functions are affine, uniform limits of affine functions are a fortiori affine. As being contiuous is also preserved by uniform limits, we conclude that $\mathcal{A}$ is uniformly closed. Now let $h \in \mathcal{A}$ and $\varepsilon > 0$. It is straightforward to check that the two sets

$$K_1 = \{(x, c) \in X \times \mathbb{R} \mid c = h(x)\}, \quad K_2 = \{(x, c) \in X \times \mathbb{R} \mid c = h(x) + \varepsilon\}$$

are compact, convex and disjoint in $E \times \mathbb{R}$, so we may apply the Hahn-Banach separation theorem to find $L \in (E \times \mathbb{R})^*$ such that

$$\sup L(K_1) < \lambda < \inf L(K_2).$$

Following the line of reasoning in Lemma 3.4, we define $g : E \to \mathbb{R}$ by the equation $L(x, g(x)) = \lambda$, which is equivalent to

$$g(x) = -\frac{L(x, 0)}{L(0, 1)} + \frac{\lambda}{L(0, 1)}.$$

This is well defined, since for any $x \in X$ we have

$$0 < L(x, g(x)) + \varepsilon) - L(x, g(x)) = L(0, \varepsilon) = \varepsilon L(0, 1),$$

from which we infer that $L(0, 1) > 0$. For any $x \in X$ we then have

$$0 < L(x, g(x)) - L(x, h(x)) = L(0, g(x) - h(x)) = L(0, 1)(g(x) - h(x)),$$

which implies $g(x) > h(x)$. The same argument yields $g(x) < h(x) + \varepsilon$. By identifying g with its restriction to X we have found $g \in E^*_{|X} + \mathbb{R}$ which satisfies $\|g - h\|_\infty < \varepsilon$, as desired. □

This proposition gives rise to a perhaps more intuitive interpretation of $\mu \sim \lambda$:

Corollary 4.3. *Let* $\mu, \lambda \in \mathcal{M}^1(X)$. *Then* $\mu \sim \lambda$ *if and only if* $r(\mu) = r(\lambda)$.

Proof. We have already remarked on the only if part. Assume then that $r(\mu) = r(\lambda) = x$. Then, for any $f + c \in E^*_{|X} + \mathbb{R}$, we have

$$\mu(f + c) = \mu(f) + c = f(x) + c = \lambda(f) + c = \lambda(f + c).$$

By the above proposition, these functions are dense in $\mathcal{A}$, and since μ and λ are continuous linear functionals on $C(X)$, we conclude that $\mu(h) = \lambda(h)$ for any $h \in \mathcal{A}$, i.e., $\mu \sim \lambda$. □

The equivalence relation provides a useful characterisation of the upper envelope:

Proposition 4.4. *Let $f \in C(X)$. Then, for every $x \in X$,*

$$\bar{f}(x) = \sup\{\mu(f) \mid \mu \sim \delta_x\}.$$

Proof. We define $g(x) = \sup\{\mu(f) \mid \mu \sim \delta_x\}$ and want to show $\bar{f} = g$. To this end we first establish that g is concave and upper semicontinuous and apply Lemma 3.4. So let $r \in \mathbb{R}$ and $\{x_\alpha\}_{\alpha \in A} \subseteq g^{-1}([r, \infty))$ be a converging net with $x = \lim_\alpha x_\alpha$. As $g(x_\alpha) \geq r$ we may for any fixed $\varepsilon > 0$ and for each $\alpha \in A$ by definition of g find $\mu_\alpha \sim \delta_{x_\alpha}$ such that

$$\mu_\alpha(f) > g(x_\alpha) - \varepsilon \geq r - \varepsilon.$$

Then $\{\mu_\alpha\}_{\alpha \in A}$ is a net in $\mathcal{M}^1(X)$, which is weak* compact by Example 2.2 and hence has a converging subnet $\{\mu_\beta\}_{\beta \in B}$. Letting $\mu = \lim_\beta \mu_\beta$, we see that for any $h \in \mathcal{A}$ we have

$$\mu(h) = \lim_\beta \mu_\beta(h) = \lim_\beta \delta_{x_\beta}(h) = \lim_\beta h(x_\beta) = h(x) = \delta_x(h).$$

Here we use the fact that $\mu_\beta \sim \delta_{x_\beta}$ for the second equality and continuity of h for the second to last equality. This establishes $\mu \sim \delta_x$, which in turn yields

$$g(x) \geq \mu(f) = \lim_\beta \mu_\beta(f) \geq r - \varepsilon$$

by choice of μ_β. Since ε was arbitrary we conclude g is upper semicontinuous. To see that g is concave we let $x_1, x_2 \in X$ and $\alpha \in (0,1)$. Then

$$\begin{aligned}\alpha g(x_1) + (1-\alpha) g(x_2) &= \sup\{\alpha\mu_1(f) + (1-\alpha)\mu_2(f) \mid \mu_1 \sim \delta_{x_1}, \mu_2 \sim \delta_{x_2}\} \\ &= \sup\{\big(\alpha\mu_1 + (1-\alpha)\mu_2\big)(f) \mid \mu_1 \sim \delta_{x_1}, \mu_2 \sim \delta_{x_2}\}.\end{aligned}$$

Since the resultant map is affine, $\mu_1 \sim \delta_{x_1}$ and $\mu_2 \sim \delta_{x_2}$ implies $\alpha\mu_1 + (1-\alpha)\mu_2 \sim \delta_{\alpha x_1 + (1-\alpha)x_2}$, and therefore the above set is a subset of the set defining $g(\alpha x_1 + (1-\alpha)x_2)$. Hence g is concave, and by applying Lemma 3.4 we see that $\bar{f} \leq g$.

To see that $g \leq \bar{f}$ let $x \in X$ and $h \in \mathcal{A}$ with $h \geq f$. We want to show that $h(x) \geq g(x)$, so let μ be any measure with $\mu \sim \delta_x$. Then, by monotonicity,

$$\mu(f) \leq \mu(h) = \delta_x(h) = h(x).$$

By definition of g this implies $g(x) \leq h(x)$, and since $\bar{f}(x)$ is the infimum over all such $h(x)$, we see that $\bar{f}(x) \geq g(x)$. □

4.2 Ordering of measures

We now define a partial order on $\mathcal{M}(X)$:

Definition 4.5. Let $\mu, \nu \in \mathcal{M}(X)$. Then $\mu \succ \nu$ if $\mu(f) \geq \nu(f)$ for all $f \in \mathcal{C}$.

The intuition behind this definition is that convex functions attain their largest values on the extreme points of X, and the partial order then measures to what degree the support of a measure is concentrated at the extreme points. The following proposition contains some basic properties of the ordering:

Proposition 4.6. *The relation $\succ$ is a partial ordering. Furthermore, it has the properties:*

1) If $\mu, \nu \in \mathcal{M}^1(X)$ and $\mu \succ \nu$, then $\mu \sim \nu$.

2) If $\mu \sim \delta_x$, then $\mu \succ \delta_x$

Proof. Since the relation is defined by inequalities, it is obviously reflexive and transitive. For antisymmetry assume $\mu \succ \nu$ and $\nu \succ \mu$. Then, for any $f - g \in \mathcal{C} - \mathcal{C}$, we have

$$\mu(f-g) = \mu(f) - \mu(g) = \nu(f) - \nu(g) = \nu(f-g).$$

Thus the two measures coincide on $\mathcal{C} - \mathcal{C}$. We shall see in the next section that $\mathcal{C} - \mathcal{C}$ is a vector lattice. It also contains the constant functions, and and since $\mathcal{A} \subseteq \mathcal{C} - \mathcal{C}$, it separates the points of X. Therefore $\mathcal{C} - \mathcal{C}$ is uniformly dense in $C(X)$ by the lattice version of the Stone-Weierstrass theorem. This is also sometimes called the Kakutani-Krein theorem, and we refer the reader to Theorem IV.12 in [10] for a proof. As μ and ν are continuous linear functionals on $C(X)$, this implies they coincide on every $f \in C(X)$, so by the uniqueness of the Riesz representation theorem, $\mu = \nu$. Hence $\succ$ is antisymmetric.

Proof of 1). Let $\mu \succ \nu$ and $h \in \mathcal{A}$. Then both h and $-h$ are convex, which yields $\mu(h) = \nu(h)$, i.e., $\mu \sim \nu$.

Proof of 2). Let $\mu \sim \delta_x$ and $f \in \mathcal{C}$. Then $-f$ is concave and continuous, which by Proposition 3.5 implies that $-f = \overline{-f}$. This, along with Proposition 4.4 applied to $-f$, implies that

$$\delta_x(-f) = -f(x) = \overline{-f}(x) \geq \mu(-f),$$

which lets us conclude that $\mu \succ \delta_x$. □

In the Choquet-Bishop-de Leeuw theorem we will produce the desired measure by choosing a measure $\mu \sim \delta_x$ which is *maximal* with respect to the above order. To do this, we need to ensure that such a measure always exists. Since this is a statement concerning maximal elements of a partial order, the key step in the proof will of course be Zorn's lemma:

Lemma 4.7. *Let $\mu \in \mathcal{M}(X)$. Then there exists a maximal $\nu \in \mathcal{M}(X)$ such that $\nu \succ \mu$.*

Proof. Regard the set $\mathcal{F} = \{\nu \in \mathcal{M}(X) \mid \nu \succ \mu\}$, partially ordered by $\succ$. We first argue why it is sufficient to find a maximal element of $\mathcal{F}$. Let ν be maximal in $\mathcal{F}$ and $\lambda \in \mathcal{M}(X)$. If $\lambda \succ \nu$ we have $\lambda \succ \mu$ by transitivity, so $\lambda \in \mathcal{F}$. Maximality then implies $\nu = \lambda$, so ν is in fact maximal in all of $\mathcal{M}(X)$.

To apply Zorn's lemma to $\mathcal{F}$, we immediately see that it is non-empty as $\mu \in \mathcal{F}$. Let $\mathcal{Z}$ be a chain in $\mathcal{F}$. Since $\nu \succ \mu$ implies that ν and μ coincide on affine functions, we have, in particular, $\nu(X) = \mu(X)$. We see that

$$\mathcal{F} \subseteq \{\nu \in \mathcal{M}(X) \mid \nu(X) = \mu(X)\} = \mu(X)\mathcal{M}^1(X).$$

Since scaling is weak* continuous, the latter is compact as the continuous image of a compact set.

We may now regard the elements of $\mathcal{Z}$ as a linearly ordered net in $\mu(X)\mathcal{M}^1(X)$, indexed by $\mathcal{Z}$ itself. This is well defined since $\mathcal{Z}$ being a chain is stronger than being a directed set. Denoting $\mathcal{Z}$ by $\{\nu_\nu\}_{\nu \in \mathcal{Z}}$, it must have a converging subnet $\{\nu_\beta\}_{\beta \in B}$ by compactness of $\mu(X)\mathcal{M}^1(X)$. Defining $\nu := \lim_\beta \nu_\beta$, we claim that this is the desired upper bound.

Indeed, let $\nu_\nu \in \mathcal{Z}$. Then we may find $\beta_0 \in B$ such that $\nu_{\beta_0} \succ \nu_\nu$. We now claim that $\nu \succ \nu_{\beta_0}$, which will prove the claim. Suppose that this is not the case, i.e., $\nu(f) < \nu_{\beta_0}(f)$ for some $f \in \mathcal{C}$. Since $\mathcal{Z}$ is linearly ordered, this means that for every $\beta \geq \beta_0$ we have $\nu_\beta \succ \nu_{\beta_0}$, so in particular $\nu_\beta(f) \geq \nu_{\beta_0}(f)$. Regarding weak* convergence as pointwise convergence, this implies

$$\nu(f) = \lim_\beta \nu_\beta(f) \geq \nu_{\beta_0}(f) > \nu(f),$$

which is a contradiction.

This establishes $\nu \succ \nu_{\beta_0} \succ \nu_\nu$ and therefore ν is an upper bound for $\mathcal{Z}$. Transitivity implies $\nu \in \mathcal{F}$, and Zorn's lemma then gives us the desired result. □

It will be a core idea in the following developments that the notion of maximality is intimately tied to the subset of X where $f(x) = \bar{f}(x)$ for *every* $f \in C(X)$. We shall see much later that this is in fact a characterisation of $\mathrm{Ext}(X)$. Therefore the following proposition relates, albeit in an indirect way, the notion of maximality to the extreme points:

Proposition 4.8. *Let $\mu \in \mathcal{M}^1(X)$ be maximal. Then $\mu(f) = \mu(\bar{f})$ for every $f \in C(X)$.*

Proof. Let $f \in C(X)$ and define the functional $T : \mathbb{R}f \to \mathbb{R}$ by $rf \mapsto r\mu(\bar{f})$. We claim that T is dominated by the sublinear functional $p : C(X) \to \mathbb{R}$ defined by $g \mapsto \mu(\bar{g})$. Proposition 3.5 ensures sublinearity of p as $\overline{rg} = r\overline{g}$ when $r \geq 0$ and $\overline{g_1 + g_2} \leq \overline{g_1} + \overline{g_2}$. This also shows that $p(rf) = T(rf)$ when $r \geq 0$. When $r < 0$ we get

$$0 = \overline{rf + (-rf)} \leq \overline{rf} + \overline{-rf},$$

which implies $-\mu(\overline{-rf}) \leq \mu(\overline{rf})$. From this we see that

$$T(rf) = -T(-rf) = -\mu(-r\bar{f}) = -\mu(\overline{-rf}) \leq \mu(\overline{rf}) = p(rf).$$

We conclude that p does indeed dominate T and immediately extend T to a linear functional $L : C(X) \to \mathbb{R}$ which is dominated by p and restricts to T. Following the usual line of reasoning, we argue that L is positive: If $g \leq 0$, we have $\bar{g} \leq 0$ since the zero function is affine continuous. This then yields

$$L(g) \leq p(g) = \mu(\bar{g}) \leq 0,$$

which implies L is non-positive on non-positive functions. As this is equivalent to being positive, we conclude L is positive linear and apply the Riesz representation theorem to produce $\nu \in \mathcal{M}(X)$ satisfying $L(g) = \int_X g \mathrm{d}\nu$ for every $g \in C(X)$. Then ν has the important property

$$\nu(f) = L(f) = T(f) = \mu(\bar{f}).$$

Furthermore, if $g \in \mathcal{C}$, $-g$ is concave and continuous, so by Proposition 3.5, $\overline{-g} = -g$. This permits the following:

$$\nu(-g) = L(-g) \leq p(-g) = \mu(\overline{-g}) = \mu(-g).$$

From this we infer that $\nu(g) \geq \mu(g)$ for all $g \in \mathcal{C}$, i.e., $\nu \succ \mu$. But due to maximality this implies $\nu = \mu$ and in particular $\mu(f) = \nu(f) = \mu(\bar{f})$, as desired. □

4.3 The Choquet-Bishop-de Leeuw theorem

In order to prove that a maximal measure is supported by every Baire set containing the extreme points of X, we have to establish a few technical lemmas. The first relies on X being metrizable, while the second generalizes the first by dropping this assumption. During the remainder of this section we surpress the index $n \to \infty$, as it otherwise clutters up the notation.

Lemma 4.9. *Let X be compact, convex and metrizable. Let $(f_n)_{n\geq 1}$ be a sequence of bounded, concave, upper semicontinuous functions on X such that the whole sequence is bounded from below and $\liminf f_n(x) \geq 0$ for every $x \in Ext(X)$. Then $\liminf f_n(x) \geq 0$ for every $x \in X$.*

Proof. Let $x_0 \in X$. By Choquet's theorem we can find a probability measure μ on X which is supported by $\mathrm{Ext}(X)$ and represents x_0. By choice of μ we now have $\liminf f_n(x) \geq 0$ μ-a.e, which implies $\mu(\liminf f_n) \geq 0$. Finding a constant $C \in \mathbb{R}$ such that $f_n \geq C$ for all $n \in \mathbb{N}$, we apply Fatou's lemma to the non-negative measurable functions $f_n - C$ and infer that

$$\mu(\liminf(f_n - C)) \leq \liminf \mu(f_n - C).$$

This and non-negativity implies

$$0 \leq \mu(\liminf f_n) \leq \liminf \mu(f_n).$$

To finish the argument we note that by Corollary 4.3, μ representing x_0 is equivalent to $\mu \sim \delta_{x_0}$. Since f_n is concave and upper semicontinuous, it equals its upper envelope. Invoking Proposition 4.4, we see that

$$f_n(x_0) = \overline{f}_n(x_0) = \sup\{\lambda(f_n) \mid \lambda \sim \delta_{x_0}\},$$

from which we infer $f_n(x_0) \geq \mu(f_n)$. Combining the pieces we find

$$\liminf f_n(x_0) \geq \liminf \mu(f_n) \geq 0.$$

Since $x_0 \in X$ was arbitrary, this yields the desired conclusion. □

We now prove the exact same lemma, removing the assumption that X is metrizable. In the proof we need that for a bounded sequence $(a_n)_{n\geq 1}$ we have $\liminf a_n = \liminf(a_n + 1/n)$. This is essentially a consequence of the fact that if $a_n \to a$, $b_n \to b$ and $a_n - b_n \to 0$, then $a = b$.

We will also need the fact that $\mathbb{R}^{\mathbb{N}}$ equipped with the product topology is a metrizable real Hausdorff LCS. Metrizability of $\mathbb{R}^{\mathbb{N}}$ is a classical theorem of point-set topology, and we refer the reader to Theorem 20.5 in [7]. It is not difficult to show that it also inherits all relevant properties from $\mathbb{R}$; one can for example verify that sets of the form $\cdots \times \mathbb{R} \times B(0,\varepsilon) \times \mathbb{R} \times \cdots$ constitute a convex neighbourhood subbase of 0.

Lemma 4.10. *Let X be compact and convex. Let $(f_n)_{n\geq 1}$ be a sequence of bounded, concave, upper semicontinuous functions on X such that the whole sequence is bounded from below and* $\liminf f_n(x) \geq 0$ *for every $x \in Ext(X)$. Then* $\liminf f_n(x) \geq 0$ *for every $x \in X$.*

Proof. Let $x_0 \in X$ and for each $n \in \mathbb{N}$ find $h_n \geq f_n$ with $h_n \in \mathcal{A}$ such that

$$h_n(x_0) < \bar{f}_n(x_0) + 1/n = f_n(x_0) + 1/n,$$

which is possible by the definition of the upper envelope. We also once again use that f_n being concave and upper semicontinuous implies $f_n = \overline{f_n}$. By the comments preceding the lemma, and since $f_n(x_0) \leq h_n(x_0) \leq f_n(x_0) + 1/n$, we see that $\liminf f_n(x_0) = \liminf h_n(x_0)$.

We then define the n'th coordinate of $\varphi : X \to \mathbb{R}^{\mathbb{N}}$ by $(\varphi(x))_n = h_n(x)$. Since each coordinate of φ is continuous and affine, so is φ itself. Therefore $\varphi(X)$ is a compact convex subset of the metrizable space $\mathbb{R}^{\mathbb{N}}$. Since each coordinate projection, $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, is linear and continuous, it is in particular bounded, concave and upper semicontinuous on $\varphi(X)$. Furthermore, for each $x \in X$,

$$f_n(x) \leq h_n(x) = \pi_n(\varphi(x)),$$

implying that boundedness from below of $(f_n)_{n\geq 1}$ is inherited by the restrictions of $(\pi_n)_{n\geq 1}$ to $\varphi(X)$. If we can justify that $\liminf \pi_n(y') \geq 0$ for each $y' \in \mathrm{Ext}(\varphi(X))$, we can apply the previous lemma to reach the desired conclusion.

To this end let $y \in \mathrm{Ext}(\varphi(X))$. By continuity of φ, the set $\varphi^{-1}(y)$ is closed in X and therefore compact. Since it is also non-empty, the classical Krein-Milman theorem establishes the existence of an extreme point $x \in \mathrm{Ext}(\varphi^{-1}(y))$. We claim that x must also be extreme in X. Indeed, letting $x = \alpha x_1 + (1-\alpha)x_2$ for $x_1, x_2 \in X$ and $\alpha \in (0,1)$, we see that

$$y = \varphi(x) = \alpha\varphi(x_2) + (1-\alpha)\varphi(x_2).$$

Since y is an extreme point of $\varphi(X)$, this implies $y = \varphi(x_1) = \varphi(x_2)$. Hence $x_1, x_2 \in \varphi^{-1}(y)$, in which x is extreme, finally yielding $x_1 = x_2 = x$. Thus $x \in \mathrm{Ext}(X)$, so by hypothesis $\liminf f_n(x) \geq 0$. Since we also have $h_n \geq f_n$, we see that

$$\liminf \pi_n(y) = \liminf \pi_n(\varphi(x)) = \liminf h_n(x) \geq \liminf f_n(x) \geq 0.$$

We now apply the previous lemma, establishing $\liminf \pi_n(y) \geq 0$ for all $y \in \varphi(X)$, not just the extreme points. Thus, for the specific x_0 from which we constructed $(h_n)_{n\geq 1}$,

$$\liminf f_n(x_0) = \liminf h_n(x_0) = \liminf \pi_n(\varphi(x_0)) \geq 0,$$

giving the conclusion. □

The two preceding lemmas allow a short proof of the following:

Lemma 4.11. *Let $(f_n)_{n\geq 1}$ be a non-decreasing sequence of functions in $C(X)$ with $-1 \leq f_n \leq 0$ and* $\lim f_n(x) = 0$ *for all $x \in Ext(X)$, and let μ be a maximal probability measure on X. Then* $\lim \mu(f_n) = 0$.

Proof. As $-1 \le f_n \le 0$, the sequence $(\overline{f}_n)_{n\ge1}$ will also be bounded by -1 and 0. Furthermore, by Proposition 3.5, each $\overline{f}_n$ is concave and upper semicontinuous. Since $\overline{f}_{n+1} \ge \overline{f}_n$ follows from $f_{n+1} \ge f_n$, $(\overline{f}_n)_{n\ge1}$ is also a bounded monotone sequence, and hence its pointwise limit exists. In particular, as $\overline{f}_n \ge f_n$, we see that for $x \in \mathrm{Ext}(X)$ we have

$$\liminf \overline{f}_n(x) = \lim \overline{f}_n(x) = \lim f_n(x) = 0,$$

the last equality stemming from the hypothesis. Appealing to the preceding lemma, we see that for any $x \in X$,

$$\lim \overline{f}_n(x) = \liminf \overline{f}_n(x) \ge 0,$$

implying that $\lim \overline{f}_n(x) = 0$. Since μ is maximal, we have $\mu(f_n) = \mu(\overline{f}_n)$ by Proposition 4.8. By boundedness of $(\overline{f}_n)_{n\ge1}$ we may then invoke the Lebesgue dominated convergence theorem to see that

$$\lim \mu(f_n) = \lim \mu(\overline{f}_n) = \mu(\lim \overline{f}_n) = 0,$$

yielding the conclusion. □

As a last stepping stone we need a technical lemma regarding regularity on Baire sets, which states that Baire sets may be approximated by G_δ and F_σ sets.

Lemma 4.12. *Let $\mu \in \mathcal{M}^1(X)$ and let B be Baire. Then for any $\varepsilon > 0$ there exists a G_δ set G and F_σ set F such that $G \subseteq B \subseteq F$ and $\mu(F \setminus G) < \varepsilon$. In particular, if $\mu(K) = 0$ for any compact $K \subseteq G \subseteq B$ where G is G_δ, then $\mu(B) = 0$.*

Proof. First define $\mathcal{E}$ to be the Borel sets which have the approximation property described above. We show that $\mathcal{E}$ is a σ-algebra which contains the compact G_δ sets.

Obviously $X \in \mathcal{E}$ since we may pick $G = X = F$. Let $A \in \mathcal{E}$ and $\varepsilon > 0$. We then find $G \subseteq A \subseteq F$ such that G is G_δ and F is F_σ and $\mu(F \setminus G) < \varepsilon$. Then $F^c \subseteq A^c \subseteq G^c$ and F^c is G_δ while G^c is F_σ. Since $F \setminus G = G^c \setminus F^c$ we have $\mu(G^c \setminus F^c) < \varepsilon$.

Now let $(A_n)_{n\ge1} \subseteq \mathcal{E}$ and let $\varepsilon > 0$. For each $n \in \mathbb{N}$ choose $G_n \subseteq A_n \subseteq F_n$ such that G_n is G_δ, F_n is F_σ and $\mu(F_n \setminus G_n) < 2^{-n}\varepsilon$. Since the countable union of countable unions of sets is once again countable, $F := \bigcup_{n\in\mathbb{N}} F_n$ is once again F_σ. Furthermore, by distributivity of $\cup$ over $\cap$, the finite union of G_δ sets is once again G_δ. Finding $N \in \mathbb{N}$ such that

$$\mu\left(\bigcup_{n=1}^{\infty} G_n \setminus \bigcup_{n=1}^{N} G_n\right) = \mu\left(\bigcup_{n=N+1}^{\infty} G_n\right) < \varepsilon,$$

we see that $G := G_1 \cup \cdots \cup G_N$ is again G_δ. Since

$$F \setminus G \subseteq \left(F \setminus \bigcup_{n=1}^{\infty} G_n\right) \cup \left(\bigcup_{n=N+1}^{\infty} G_n\right) \subseteq \left(\bigcup_{n=1}^{\infty} F_n \setminus G_n\right) \cup \left(\bigcup_{n=N+1}^{\infty} G_n\right),$$

we see that

$$\mu(F \setminus G) \le \sum_{n=1}^{\infty} \mu(F_n \setminus G_n) + \mu\left(\bigcup_{n=N+1}^{\infty} G_n\right) \le 2\varepsilon.$$

Since $G \subseteq \bigcup_{n\in\mathbb{N}} A_n \subseteq F$, we see that $\mathcal{E}$ is a σ-algebra. Now, if G is compact G_δ, it is also closed and therefore F_σ, so obivously $G \in \mathcal{E}$. Since the compact G_δ sets generate the Baire σ-algebra, all Baire sets are in $\mathcal{E}$. The last statement of the lemma is then a simple combination of this approximation property and inner regularity of μ. □

We have now paved the way for a relatively short proof of this chapter's main theorem:

Theorem 4.13 (Choquet-Bishop-de Leeuw)**.** *Let X be a compact and convex subset of E. For every $x \in X$ there exists a probability measure μ on X which represents x and is supported by every Baire set containing the extreme points of X.*

Proof. Let $x_0 \in X$ and let μ be a maximal probability measure with $\mu \sim \delta_{x_0}$, the existence of which is the statement of Lemma 4.7. By Corollary 4.3 this implies μ represents x_0, so we need only show that μ is supported by every Baire set containing $\mathrm{Ext}(X)$. If B is such a Baire set, let $G \subseteq B^c$ be G_δ and let $K \subseteq G$ be compact. Then G^c is F_σ, so

$$\mathrm{Ext}(X) \subseteq G^c = \bigcup_{k=1}^{\infty} F_k$$

for some closed sets F_k. Then, for any $n \in \mathbb{N}$, the set $C_n := \bigcup_{k=1}^{n} F_k$ is closed and disjoint from K. Since X is compact Hausdorff and translation by -1 is continuous, we may apply Urysohn's lemma to find $f_n : X \to [-1, 0]$ such that $f_n(K) = -1$ and $f_n(C_n) = 0$ for all $n \in \mathbb{N}$. Defining $g_1 = f_1$ and $g_n = \max\{f_n, g_{n-1}\}$, the sequence $(g_n)_{n\geq 1}$ is continuous and non-decreasing on X and still satisfies $g_n(K) = -1$ and $g_n(C_n) = 0$ for each $n \in \mathbb{N}$. Furthermore, for $x \in \mathrm{Ext}(X)$ we eventually have $x \in C_n$, implying $\lim g_n(x) = 0$.

We are now in the situation of Lemma 4.11 which lets us conclude $\lim \mu(g_n) = 0$. Assuming $\mu(K) > 0$, we see that

$$\mu(g_n) = \int_X g_n \mathrm{d}\mu = \int_K g_n \mathrm{d}\mu + \int_{K^c} g_n \mathrm{d}\mu = -\mu(K) + \int_{K^c} g_n \mathrm{d}\mu \leq -\mu(K),$$

using that $g_n \leq 0$ for the last inequality. This violates $\lim \mu(g_n) = 0$, so we conclude $\mu(K) = 0$. Then, by the preceding lemma, μ is indeed supported by every Baire set containing $\mathrm{Ext}(X)$. $\square$

This concludes our survey of theorems regarding the existence of representing measures supported by the extreme points. In many ways Choquet's theorem provided exactly what we wanted; representing measures supported by just the extreme points. But this relied heavily on the existence of a strictly convex function and the extreme points being Baire, properties which are not guaranteed in the general situation. The Choquet-Bishop-de Leeuw theorem then improves the reformulated Krein-Milman in the sense that the produced representing measure is supported by every Baire set containing the extreme points, not just its closure.

We saw in Example 2.10 that being supported by the closure of the extreme points can be very different from being supported by the extreme points themselves. In the same vein there exist pathological examples of measures supported by every Baire set containing the extreme points which still measure the extreme points themselves to 0. One such example is due to Mokobodzki and is detailed on p. 60 in [8].

5 Vector lattices and Choquet simplices

In this section we will introduce the concepts of lattices and Choquet simplices, as these are the natural framework for discussing uniqueness of representing measures. The material in this and the next section is based on the content of chapter 10 in [8]. We start by defining vector lattices in general and then turn to applying the concept in the setting of compact convex sets in a locally convex space. We begin with the definition:

Definition 5.1. A partially ordered real vector space $(V, \leq)$ is a *vector lattice* if for all $x, y, z \in V$ the following conditions hold:

1. If $x \leq y$, then $x + z \leq y + z$.
2. If $x \leq y$ and $\alpha \geq 0$, then $\alpha x \geq \alpha y$.
3. For any $x, y \in V$ there exists a least upper bound denoted by $x \vee y$.

We will usually refer to the first property as translation invariance and to the second as positive homogeneity, while $x \vee y$ is called the *join* of x and y. We shall see that any $x, y \in V$ also have a greatest lower bound $x \wedge y$, called the *meet* of x and y. Vector lattices are also often referred to as *Riesz spaces* in the literature. As is commonplace after defining a mathematical structure, there are a handful of easily obtained properties which follow in a straightforward manner from the definition. We record those that will be useful to us:

Proposition 5.2. *Let V be a vector lattice and let $x, y, z, w \in V$.*

1. *If $x \leq y$ and $z \leq w$, then $x + z \leq y + w$. Furthermore, $x \geq y \Leftrightarrow -x \leq -y$.*
2. *$x \wedge y = -(-x \vee -y)$ and $x \vee y = -(-x \wedge -y)$.*
3. *$(x \wedge y) + z = (x + z) \wedge (y + z)$ and $(x \vee y) + z = (x + z) \vee (y + z)$.*
4. *If $x \leq 0$, $y \leq 0$ and $z \leq 0$, then $(x + y) \wedge z \leq (x \wedge z) + (y \wedge z)$.*

Proof. The first two statements follow immediately from the definition. Let us prove the third. As $x \wedge y \leq x$ and $x \wedge y \leq y$, we get $(x \wedge y) + z \leq x + z$ and $(x \wedge y) + z \leq y + z$ by translation invariance, so $(x \wedge y) + z$ is in fact a lower bound for $(x + z)$ and $(y + z)$.

Now let w be a lower bound for $(x + z)$ and $(y + z)$. Then $w \leq x + z$, which means $w - z \leq x$. Similarly $w - z \leq y$, so $w - z \leq x \wedge y$. Thus $w \leq (x \wedge y) + z$ as desired. The second statement is similar.

To prove *4.* let $u := (x + y) \wedge z$. Then $u \leq x + y$ and $u \leq z$ as it is the greatest lower bound of the two. The latter implies, together with $0 \leq x$, that $0 + u = u \leq x + z$. This however, together with $u \leq x + y$, gives us

$$u \leq (x + y) \wedge (x + z) = x + (y \wedge z)$$

by part *3.* As both y and z are non-negative, $0 \leq y \wedge z$. Combining this with $u \leq z$, we see that $u \leq z + (y \wedge z)$. Using the two inequalities and distributivity of $+$ once again, we see that

$$u \leq (x + (y \wedge z)) \wedge (z + (y \wedge z)) = (x \wedge z) + (y \wedge z),$$

finishing the proof. □

We notice for later use that the second statement implies that one may equivalently check the greatest lower bound property when verifying that a given object is a vector lattice. We have previously used the fact that the subspace $\mathcal{C} - \mathcal{C}$ is a vector lattice, so it is about time that we prove it. For this we need the following lemma:

Lemma 5.3. *If $f, g \in \mathcal{C}$, then $\max(f, g) \in \mathcal{C}$. Similarly, if $f, g \in -\mathcal{C}$, then $\min(f, g) \in -\mathcal{C}$.*

Since the proof is quite straightforward, we omit it.

Proposition 5.4. *The set $\mathcal{C} - \mathcal{C} = \{f - g \mid f, g \in \mathcal{C}\}$ is a vector lattice under pointwise ordering and pointwise maximum as the join operation.*

Proof. Convexity is stable under addition and positive scaling, implying that $\mathcal{C} - \mathcal{C}$ is a subspace of $C(X)$. Furthermore, if $f, g \in \mathcal{C}$, then $\max(f, g) \in \mathcal{C}$ by the preceding lemma. The conclusion then follows from the fact that for $f_1 - g_1, g_2 - g_2 \in \mathcal{C} - \mathcal{C}$ we have

$$\max(f_1 - g_1, f_2 - g_2) = \max(f_1 + g_2, f_2 + g_1) - (g_1 + g_2).$$

This is most easily seen by distinguishing between two entirely symmetric cases. For a fixed $x \in X$ assume $(f_1 - g_1)(x) \geq (f_2 - g_2)(x)$, i.e., $f_1(x) + g_2(x) \geq f_2(x) + g_1(x)$. The right hand side of the above equation then reads

$$\max(f_1 + g_2, f_2 + g_1)(x) - (g_1 + g_2)(x) = f_1(x) - g_1(x),$$

which is exactly what the left hans side equates to. Similiarly the identity also holds when $(f_1 - g_1)(x) \leq (f_2 - g_2)(x)$.

Since $\max(f_1 + g_2, f_2 + g_1)$ and $g_1 + g_2$ are elements of $\mathcal{C}$, $\max(f_1 - g_1, f_2 - g_2) \in \mathcal{C} - \mathcal{C}$, and therefore $\mathcal{C} - \mathcal{C}$ does indeed have the least upper bound property. Since translation invariance and positive homogeneity are immediate, we conclude that $\mathcal{C} - \mathcal{C}$ is in fact a vector lattice. □

5.1 The Decomposition lemma

There is one lemma from lattice theory which will be absolutely crucial in proving the uniqueness theorem for representing measures. This is the so called *Decomposition lemma*:

Lemma 5.5 (Decomposition lemma)**.** *Let V be a vector lattice and let $\{x_i \mid i \in I\}$ and $\{y_j \mid j \in J\}$ be finite sets of non-negative elements of V such that $\sum_{i \in I} x_i = \sum_{j \in J} y_j$. Then there exists a set $\{z_{ij} \in V \mid (i, j) \in I \times J\}$ with $z_{ij} \geq 0$ such that for every $i \in I$ and $j \in J$ we have*

$$x_i = \sum_{j \in J} z_{ij} \qquad \text{and} \qquad y_j = \sum_{i \in I} z_{ij}.$$

Proof. One quickly realizes that there is little to show when either $|I| = 1$ or $|J| = 1$. We show that the lemma holds for $|I| = |J| = 2$ and reduce the general statement to this by induction. We assume without loss of generality that $I = J = \{1, 2\}$. Let $x_i \geq 0$ and $y_j \geq 0$ for $i, j \in \{1, 2\}$ with $x_1 + x_2 = y_1 + y_2$. To prove the statement we seek elements $z_{ij} \geq 0$ to fill out the matrix

$$\begin{array}{cc|c} z_{11} & z_{12} & x_1 \\ z_{21} & z_{22} & x_2 \\ \hline y_1 & y_2 & S \end{array}$$

where S denotes the common sum and each x_i is the sum of the row it appears in and similarly each y_j is the sum of the column it appears in.

We start by defining $z_{11} = x_1 \wedge y_1$, which is non-negative as x_1 and y_1 are non-negative. This forces us to define $z_{12} = x_1 - z_{11}$ and $z_{21} = y_1 - z_{11}$, which by construction are also both non-negative. Finally let $z_{22} = x_2 - z_{21}$. For the lemma to hold, this must equal $y_2 - z_{12}$. Indeed, inserting the different definitions and using that $y_1 = x_1 + x_2 - y_2$, we have

$$x_2 - z_{21} = x_2 - y_1 + z_{11} = x_2 - (x_1 + x_2 - y_2) + z_{11} = y_2 - x_1 + z_{11} = y_2 - z_{12}.$$

Then $x_2 = z_{21} + z_{22}$ and $y_2 = z_{12} + z_{22}$, so all the required identities hold. We still need to argue that $z_{22} \geq 0$. Since $z_{21} \leq z_{21}$ and $0 \leq y_2$ we have $z_{21} \leq z_{21} + y_2 = x_2 + z_{12}$. Thus z_{21} is a lower bound for z_{21} and $x_2 + z_{12}$. Now if u is some other lower bound for the two we obviously have $u \leq z_{21}$, from which we infer that $z_{21} = z_{21} \wedge (x_2 + z_{12})$. By definition of z_{21} and z_{12} we have

$$z_{21} \wedge z_{12} = (x_1 - z_{11}) \wedge (y_1 - z_{11}) = (x_1 \wedge y_1) - z_{11} = 0,$$

using distributivity of $+$. Putting the different pieces together and using part *4.* of Proposition 5.2 yields

$$z_{21} = z_{21} \wedge (z_{12} + x_2) \leq (z_{21} \wedge z_{12}) + (z_{21} \wedge x_2) = z_{21} \wedge x_2.$$

From this we see that $z_{21} \leq x_2$, and therefore $z_{22} = x_2 - z_{21} \geq 0$.

Thus we have proven the lemma when $|I| = |J| = 2$. Let us generalise to the case when $|I| = n$ and $|J| = 2$, which is actually the case we will be using later. We do this by induction, so let $n > 2$ and assume the lemma is true for $n' \leq n - 1$. Let $\{x_i \mid i \in I\}$ and y_1, y_2 be non-negative such that

$$\sum_{i=1}^{n} x_i = y_1 + y_2 = (x_1 + x_2) + \sum_{i=3}^{n} x_i$$

By the induction hypothesis we may then find $z_{ij} \geq 0$ such that we get the sums indicated in the following matrix:

$$\begin{array}{cc|c} z_{11} & z_{12} & x_1 + x_2 \\ \vdots & \vdots & \vdots \\ z_{(n-1,1)} & z_{(n-1,2)} & x_n \\ \hline y_1 & y_2 & S \end{array}$$

The first row implies $z_{11} + z_{12} = x_1 + x_2$, so we may use the base case shown earlier to find $w_{ij} \geq 0$ and the sums represented by the matrix

$$\begin{array}{cc|c} w_{11} & w_{12} & x_1 \\ w_{21} & w_{22} & x_2 \\ \hline z_{11} & z_{12} & \end{array}$$

As a final step we concatenate the two matrices to get the desired non-negative elements and sums:

$$\begin{array}{cc|c} w_{11} & w_{12} & x_1 \\ w_{21} & w_{22} & x_2 \\ z_{21} & z_{22} & x_3 \\ \vdots & \vdots & \vdots \\ z_{(n-1,1)} & z_{(n-1,2)} & x_n \\ \hline y_1 & y_2 & S \end{array}$$

To finish the argument one applies the exact same idea to the case when $m \in \mathbb{N}$ is arbitrary. The only difference is that it is even worse notationally! □

5.2 Cones and simplices

After this rather technical lemma we return to some definitions, namely those of a cone and of a Choquet simplex.

Definition 5.6. Let V a real vector space. A non-empty subset $P \subseteq V$ is a *convex cone* if $x, y \in P$ and $\alpha, \beta \geq 0$ implies $\alpha x + \beta y \in P$. A convex set $B \subseteq P$ is a *base* for P if $B = H \cap P$ for a closed hyperplane H missing the origin and B generates P, that is, $\tilde{B} := \{\alpha x \mid \alpha \geq 0, x \in B\} = P$.

We can always find a continuous linear functional such that $H = L^{-1}(1)$. From this it follows that the representation of non-zero elements of P is unique, i.e., if $\alpha_1 x_1 = \alpha_2 x_2$ for $\alpha > 0$ and $x_1, x_2 \in B$ we have $\alpha_1 = \alpha_2$ and $x_1 = x_2$. Since H misses the origin, $\alpha x = 0$ at least implies $\alpha = 0$. We will use this very frequently in the following developments.

A convex cone with base B gives rise to a partial order on E, as we shall see in the following proposition:

Proposition 5.7. *Let V be a real vector space and $P \subseteq V$ a cone with base B. Then P induces a translation invariant and homogenous partial order $\geq$ on V by $x \geq y \Leftrightarrow x - y \in P$.*

Proof. Reflexivity is obvious. For antisymmetry it suffices to show that $P \cap (-P) = \{0\}$. If $y \in P \cap (-P)$, then $y = \alpha_1 x_1 = -\alpha_2 x_2$ for some $\alpha_1, \alpha_2 \geq 0$ and $x_1, x_2 \in B$. Assuming $\alpha_2 \neq 0$, we get $P \ni x_2 = -\frac{\alpha_1}{\alpha_2} x_1$. From the uniqueness of representation discussed earlier, it follows that $1 = -\frac{\alpha_1}{\alpha_2}$, a contradiction. Hence $\alpha_2 = 0$ and thus $y = 0$. Transitivity, translation invariance and homogeneity are also immediate. □

Since we are interested in studying *any* compact convex subset of a locally convex space E, it is an important question to ask whether we can always find a cone P such that X is its base. While this is not true in general, it is true to the extent we care about: X is affinely homeomorphic to $X \times \{1\} \subseteq E \times \mathbb{R}$ through the map $\varphi(x) = (x, 1)$. Then $\varphi(X)$ is a compact convex set contained in a hyperplane missing the origin in the LCS $E \times \mathbb{R}$, and hence $\varphi(X)$ is a base for $\widetilde{\varphi(X)}$. To avoid carrying around a homeomorphism and a trivial second coordinate, we will simply assume without loss of generality that X is contained in a closed hyperplane missing the origin. For such an X it is straightforward to check that $\tilde{X}$ is indeed a convex cone with base X.

Next we define the notion of a Choquet simplex:

Definition 5.8. Let B be a base for a convex cone P. Then B is a *Choquet simplex* if $P - P$ is a vector lattice in the ordering induced by P.

This is most definitely not the usual definition of a simplex and does not immediately lend itself to any intuition about why sets satisfying this should resemble triangles. We shall, however, see in the sequel that it is a quite handy definition which extends well the finite-dimensional notion to the general setting. Note that $P - P$ is a subspace and is equipped with the translation invariant homogenous partial order induced by P. Thus, when verifying that B is a Choquet simplex, the only criteria to be checked is the least upper (or greatest lower) bound property of $P - P$.

Since we are only interested in properties of compact convex sets invariant under affine homeomorphisms, we argue that being a Choquet simplex is an intrinsic property:

Proposition 5.9. *Let $B \subseteq E$, $B' \subseteq E'$ be convex sets and let $\varphi : B \to B'$ be an affine bijection. Then φ extends to an additive and order-preserving map $\bar{\varphi} : \tilde{B} \to \tilde{B}'$. In particular, B is a Choquet simplex if and only if B' is.*

Proof. We naturally define $\bar{\varphi}(\alpha x) = \alpha \varphi(x)$, which is well-defined by the uniqueness of representation. To see that $\bar{\varphi}$ additive, let $\alpha_1 x_1, \alpha_2 x_2 \in \tilde{B}$ with one α_i non-zero. We then have

$$\begin{aligned}
\bar{\varphi}(\alpha_1 x_1 + \alpha_2 x_2) &= \bar{\varphi}\left((\alpha_1 + \alpha_2)\left(\frac{\alpha_1}{\alpha_1 + \alpha_2} x_1 + \frac{\alpha_2}{\alpha_1 + \alpha_2} x_2\right)\right) \\
&= (\alpha_1 + \alpha_2)\left(\frac{\alpha_1}{\alpha_1 + \alpha_2}\varphi(x_1) + \frac{\alpha_2}{\alpha_1 + \alpha_2}\varphi(x_2)\right) \\
&= \alpha_1 \varphi(x_1) + \alpha_2 \varphi(x_2) = \bar{\varphi}(\alpha_1 x_1) + \bar{\varphi}(\alpha_2 x_2).
\end{aligned}$$

Here we used φ being affine in the second line. It then follows in a straightforward manner from the definitions that $\bar{\varphi}$ is order-preserving and that $\tilde{B}'$ has the least upper bound property if $\tilde{B}$ does. Since the inverse of an affine bijection is once again affine, the converse also holds. □

While it is often more convenient to work with the vector lattice $P - P$, it is also sometimes useful to work with just P. The content of the next proposition is that one of these has the least upper bound property if and only if the other does, i.e., $P - P$ is a vector lattice if and only if P is a lattice:

Proposition 5.10. *Let B be a base for a convex cone P. In the ordering induced by P, $P - P$ is a vector lattice if and only if P is a lattice.*

Proof. If $P - P$ is a lattice and $x, y \in P$, we only need to show that $x \vee y \in P$. But as $x \vee y \geq x$, we see that $(x \vee y) - x = \beta z$ for some $\beta \geq 0$, $z \in X$. Writing $x = \alpha x'$, we see that $x \vee y = \beta z + \alpha x' \in P$.

Assume P is a lattice and let $x = x_1 - x_2, y = y_1 - y_2$ be arbitrary elements of $P - P$. Define

$$z = ((x_1 + y_2) \vee (y_1 + x_2)) - (x_2 + y_2).$$

We then see that

$$z - x = ((x_1 + y_2) \vee (y_1 + x_2)) - (x_2 + y_2) - (x_1 - x_2) = ((x_1 + y_2) \vee (y_1 + x_2)) - (x_1 + y_2),$$

which is an element of P and thus $z \geq x$. Similiary $z \geq y$, so z is an upper bound for x and y. Next let $w \in P - P$ be another upper bound for the two. Writing $w = w_1 - w_2$, we see that $w \geq x$ implies that $w_1 + x_2 \geq x_1 + w_2$. Using translation invariance yields $w_1 + x_2 + y_2 \geq w_2 + x_1 + y_2$. By swapping the roles of x and y we get the symmetric statement $w_1 + y_2 + x_2 \geq w_2 + y_1 + x_2$. Having these two in mind we do the following calculation:

$$\begin{aligned} w - z &= w_1 - w_2 - ((x_1 + y_2) \vee (y_1 + x_2) - (x_2 + y_2)) \\ &= w_1 + x_2 + y_2 - (w_2 + ((x_1 + y_2) \vee (y_1 + x_2)) \\ &= w_1 + x_2 + y_2 - ((w_2 + x_1 + y_2) \vee (w_2 + y_1 + x_2)) \in P. \end{aligned}$$

Here we used distributivity of + and the two prior inequalities. This implies $w \geq z$ and hence z is in fact a least upper bound in $P - P$. □

The above proposition lends itself to a nice geometrical interpretation of being a Choquet simplex. To see this, note that for $x \in P$ we have $\{z \in P \mid z \geq x\} = x + P$. Therefore, if $z = x \vee y$, we must have $(x + P) \cap (y + P) = z + P$. And conversely, if $(x + P) \cap (y + P) = z + P$ for some $z \in P$, it is easy to verify that z must be the least upper bound of x and y. Therefore P is a lattice exactly if the intersection of any two translated cones in P is again a translated cone. The image from p. 159 in [1] illustrates extremely clearly how a circle is *not* a Choquet simlex, while a triangle is.

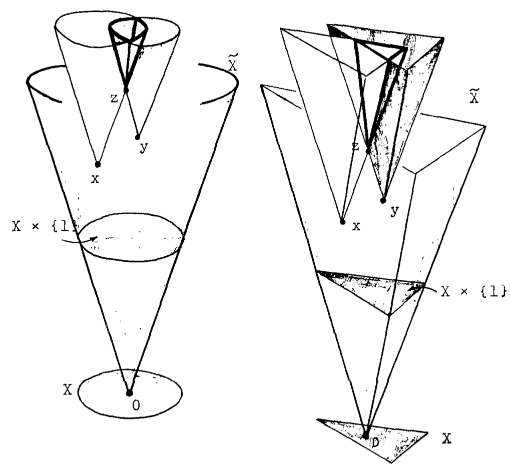

Figure 2: From Figure 28.3 in *Lectures on Analysis* by G. Choquet

In the proof of uniqueness of maximal representing measures it will be crucial that the set of maximal probability measures is a lattice. For this we define the notion of a hereditary subcone:

Definition 5.11. Let V be a real vector space, let $P_1, P_2 \subseteq V$ be convex cones with $P_1 \subseteq P_2$ and denote the two induced orderings by $\leq_1$, $\leq_2$, respectively. P_1 is a hereditary subcone of P_2 if $x \in P_1$, $y \in P_2$ and $y \leq_2 x$ implies $y \in P_1$.

While the notion of a hereditary subcone may be difficult to visualise, we shall see that the maximal measures form a hereditary subcone of all measures. We prove the following short lemma:

Lemma 5.12. *Let V be a real vector space and let P_1 be a hereditary subcone of P_2. If P_2 is a lattice in its own ordering, P_1 is also a lattice in its own ordering.*

Proof. Let $x, y \in P_1$ and define $z = x \wedge_2 y$, where $\wedge_2$ denotes the meet in P_2. Then $z \leq_2 x$, so by the definition of a hereditary subcone $z \in P_1$. Since $x - z = w$ for some $w \in P_2$, we see that $x - w = z \in P_2$, so $x \geq_2 w$, which once again implies that $w \in P_1$. From this we obtain $z \leq_1 x$ and similarly $z \leq_1 y$, so z is in fact a lower bound in the ordering induced by P_1.

Now let $w \in P_1$ be a lower bound for x and y. Then, as $P_1 \subseteq P_2$, $w \leq_2 x$ and $w \leq_2 y$, from which we gather $w \leq_2 z$ or $z - w \in P_2$. Futhermore, as $w \in P_2$, we have $0 \leq_2 w$, and hence $0 \geq_2 -w$. From this we infer that $z - w \leq_2 z$. We argued earlier that $z \in P_1$, so by the property of being hereditary, we see that $z - w \in P_1$. From this we obtain $z \geq_1 w$ as desired. □

Before we apply the lemma to maximal measures, we shall see our first example of a Choquet simplex:

Proposition 5.13. *The set of probability measures $\mathcal{M}^1(X)$ is a Choquet simplex.*

Proof. We have already seen that $\mathcal{M}^1(X)$ is compact and convex in the weak* topology of $C(X)^*$ and contained in the closed hyperplane $\hat{1}^{-1}(1)$. Furthermore, emphasizing the interpretation of Radon measures as positive linear functionals, $\mathcal{M}(X)$ is easily seen to be a convex cone generated by $\mathcal{M}^1(X)$. Using Proposition 5.10, we then show that $\mathcal{M}(X)$ is a lattice in its own ordering. We note that for $\rho, \mu \in \mathcal{M}(X)$ we have

$$\rho \leq \mu \quad \Leftrightarrow \quad \mu - \rho \in \mathcal{M}(X) \quad \Leftrightarrow \quad \mu - \rho \text{ positive linear.}$$

We will also use that if $f_1 \geq f_2$ are non-negative measurable, then $f_1 \cdot \mu - f_2 \cdot \mu = (f_1 - f_2) \cdot \mu$. This follows from

$$f_1 \cdot \mu = ((f_1 - f_2) + f_2) \cdot \mu = (f_1 - f_2)\mu + f_2\mu.$$

Now let $\mu, \lambda \in \mathcal{M}(X)$. Defining $\nu = \mu + \lambda$, we see that both μ and λ are absolutely continuous with respect to ν. By the Lebesgue-Radon-Nikodym theorem (Theorem 3.8 in [3]), there exist non-negative measurable functions f_1 and f_2 such that $\mu = f_1 \cdot \nu$ and $\lambda = f_2 \cdot \nu$. Defining $g = \min(f_1, f_2)$, we claim that $\mu \wedge \lambda = g \cdot \nu$. In the following we will repeatedly and silently appeal to Proposition 1.9 to know that the measures constructed are in fact Radon.

We see that $g \cdot \nu$ is a lower bound for μ, as

$$\mu - g \cdot \nu = f_1 \cdot \nu - g \cdot \nu = (f_1 - g) \cdot \nu \in \mathcal{M}(X).$$

Substituting λ for μ of course yields $g \cdot \nu \leq \lambda$. Hence $g \cdot \nu$ is a lower bound for μ and λ.

Now let ρ be a lower bound for μ and λ. Then $\mu - \rho \in \mathcal{M}(X)$, which implies ρ is absolutely continuous w.r.t μ and hence also w.r.t ν. Thus $\rho = f_3 \cdot \nu$ for some non-negative measurable f_3. Then

$$\mu - \rho = f_1 \cdot \nu - f_3 \cdot \nu \in \mathcal{M}(X).$$

From this it follows that $f_1 \geq f_3$ ν-a.e., so we might as well assume $f_1 \geq f_3$, since the f_i's are only defined ν-a.e to begin with. Swapping the roles of μ and λ yields $f_2 \geq f_3$, which implies $f_3 \leq g$. This in turn yields $\rho \leq g \cdot \nu$ as desired. □

When proving that having unique maximal measures implies being a Choquet simplex, it will be of essence that the maximal probability measures are a Choquet simplex themselves. We conclude this section by proving this fact, using the lemma concerning hereditary subcones:

Lemma 5.14. *Let $\mathcal{Q} := \{\mu \in \mathcal{M}(X) \mid \mu \text{ maximal}\}$ and let $\mathcal{Q}^1$ be the set of maximal probability measures. Then $\mathcal{Q}$ is a subcone of $\mathcal{M}(X)$ with base $\mathcal{Q}^1$. Furthermore, $\mathcal{Q}^1$ is a Choquet simplex.*

Proof. We first show that $\mathcal{Q}$ is a subcone. Letting $\mu, \lambda \in \mathcal{Q}$, $\alpha, \beta \geq 0$ and $f \in C(X)$, we see that

$$\alpha\mu(f) + \beta\lambda(f) = \alpha\mu(\bar{f}) + \beta\lambda(\bar{f}) = (\alpha\mu + \beta\lambda)(\bar{f}).$$

This is actually equivalent to $\alpha\mu + \beta\lambda \in \mathcal{Q}$, the proof of which we delay until Proposition 6.3 in the next section. This also shows that $\mu/\mu(1)$ is maximal, so $\mathcal{Q} = \tilde{\mathcal{Q}}^1$. Furthermore, $\mathcal{Q}^1$ is convex and contained in $\hat{1}^{-1}(1)$, so we conclude that $\tilde{\mathcal{Q}}^1$ is in fact a base for $\mathcal{Q}$.

To show that $\mathcal{Q}^1$ is a Choquet simplex, we show that $\mathcal{Q}$ is a lattice in its own ordering. We just proved that $\mathcal{M}^1(X)$ is a Choquet simplex, and therefore by Lemma 5.12, it suffices to establish that $\mathcal{Q}$ is a hereditary subcone of $\mathcal{M}(X)$. To this end let $\mu \in \mathcal{Q}$, $\lambda \in \mathcal{M}(X)$ and $\lambda \leq \mu$ in the ordering induced by $\mathcal{M}(X)$, i.e., $\mu - \lambda \in \mathcal{M}(X)$. We claim that λ is maximal. Indeed, if $\nu \succ \lambda$ we get

$$\nu + (\mu - \lambda) \succ \lambda + (\mu - \lambda) = \mu.$$

Along with using $\mu - \lambda \in \mathcal{M}(X)$ we have also used the fact that $\succ$ is stable under addition, which is obvious from the definition. Then μ being maximal implies $\nu + (\mu - \lambda) = \mu$, from which we get $\nu = \lambda$ as desired. Hence $\lambda \in \mathcal{Q}$, so $\mathcal{Q}^1$ is a Choquet simplex. $\square$

6 Uniqueness

In this section we prove the long awaited uniqueness theorem for representing measures, which states that maximal representing measures are unique if and only if the compact convex set of interest is a Choquet simplex. Before we plunge into the proof of the theorem itself, we prove a number of lemmas and propositions which will help us later. Many of these are refinements or converses of earlier results, so one may regard the next couple of pages as a sharpening of the tools already at hand.

6.1 Maximality revisited

The purpose of this section is to give a very useful equivalent characterisation of maximality using upper envelopes. We first establish a couple of auxiliary results. To moderate the length of individual proofs, we begin with this purely measure theoretic lemma:

Lemma 6.1. *Let $\mu \in \mathcal{M}(X)$ and $f \in L^1$ with $f(x) = \inf\{g(x) \mid g \in G\}$ for some $G \subseteq C(X)$ which is directed downwards, i.e., for each $g_1, g_2 \in G$ there exists $g_3 \in G$ such that $g_3 \leq g_1$ and $g_3 \leq g_2$ in the pointwise ordering. We then have*

$$\mu(f) = \inf\{\mu(g) \mid g \in G\}.$$

Proof. Defining $\beta = \inf\{\mu(g) \mid g \in G\}$ we seek to prove $\mu(f) = \beta$. To this end we find a sequence $(g'_n)_{n\geq 1} \subseteq G$ for which we have $\lim_{n\to\infty} \mu(g'_n) = \beta$. Letting $\mathfrak{L}(g_1, g_2)$ denote the set of lower bounds for g_1 and g_2 in G, we define $g_1 = g'_1$ and choose $g_n \in \mathfrak{L}(g'_n, g_{n-1})$ for all $n \in \mathbb{N}$, invoking the axiom of choice. Notice that we can't simply choose $\min(g'_n, g_{n-1})$, since we have not assumed G to be stable under minimums. Then $(g_n)_{n\geq 1} \subseteq G$ is monotonically decreasing while monotonicity of the integral ensures that $\lim_{n\to\infty} \mu(g_n) = \beta$.

By hypothesis $g_n(x) \geq f(x)$ for all $x \in X$ and $n \in \mathbb{N}$. Thus $(g_n)_{n\geq 1}$ is bounded from below, so it will converge pointwise to a measurable function g for which it holds that $g_1 \geq g \geq f$. Then $f, g_1 \in L^1$ implies $g \in L^1$. By applying the monotone convergence theorem we then obtain

$$\mu(g) = \lim_{n\to\infty} \mu(g_n) = \beta.$$

From $g \geq f$ and monotonicity we conclude $\beta \geq \mu(f)$.

Letting $A = \{x \in X \mid g(x) > f(x)\}$ it sufficient to prove $\mu(A) = 0$ to infer $\beta = \mu(g) = \mu(f)$. Assume that this is not the case, i.e., $\mu(A) > 0$. We may then find $r \in \mathbb{R}$, $\varepsilon > 0$ such that $B := \{x \in X \mid g(x) > r, f(x) < r - \varepsilon\}$ has positive measure. Indeed, if this were not the case, we would have

$$\mu\left(\bigcup_{q\in\mathbb{Q}} \bigcup_{n\in\mathbb{N}} \{x \in X \mid g(x) > q, f(x) < q - 1/n\}\right) = 0.$$

But this set contains A, which under present assumptions has positive measure. Thus $\mu(B) > 0$.

We noted in the preliminaries that Radon probability measures are in fact inner regular on all Borel sets, not just open ones. Hence there exists a compact set $K \subseteq B$ such that $\mu(K) > 0$. For each $x \in K$ we have $f(x) < r - \varepsilon$, and since $f(x) = \inf\{h(x) \mid h \in G\}$, we may find $h_x \in G$ such that $h_x(x) < r - \varepsilon$. Then

$$K \subseteq \bigcup_{x\in K} h_x^{-1}((-\infty, r - \varepsilon)),$$

which by continuity of h_x is an open cover. Finding $h_1, \ldots, h_n$ which yield a finite subcover, we define a sequence $(f_k)_{k\geq 1}$ by choosing $f_k \in \mathfrak{L}(g_k, h_1, \ldots, h_n)$. We see that $(f_k)_{k\geq 1} \subseteq G$ and $\lim_{k\to\infty} \mu(f_k) = \beta$ by monotonicity. Now, for $x \in K$ we have

$$f_k(x) < r - \varepsilon < g(x) - \varepsilon \leq g_k(x) - \varepsilon,$$

where the first inequality follows from the fact that $x \in h_i^{-1}((\infty, r-\varepsilon))$ for some i while the second is a consequence of $K \subseteq B$. For $x \in X \setminus K$ we furthermore have $f_k(x) \leq g_k(x)$. We may now reap the harvest of our hard work, since this yields

$$\beta \leq \mu(f_k) = \int_K f_k \mathrm{d}\mu + \int_{X\setminus K} f_k \mathrm{d}\mu \leq \int_K g_k - \varepsilon \mathrm{d}\mu + \int_{X\setminus K} g_k \mathrm{d}\mu = \mu(g_k) - \mu(K)\varepsilon.$$

Hence $\beta < \beta + \mu(K)\varepsilon \leq \lim_{k\to\infty} \mu(g_k) = \beta$, a contradiction. $\square$

This lemma helps us establish a different perspective on the upper envelope:

Lemma 6.2. *Let $f \in C(X)$ and define $G_f = \{g \in -\mathcal{C} \mid g > f\}$. Then $\bar{f}(x) = \inf\{g(x) \mid g \in G_f\}$ for each $x \in X$. Furthermore, G_f is directed downwards by pointwise minimum and $\mu(\bar{f}) = \inf\{\mu(g) \mid g \in G_f\}$ for each $\mu \in \mathcal{M}(X)$.*

Proof. Let $x \in X$ and define $f'(x) = \inf\{g(x) \mid g \in G_f\}$. Since $\mathcal{A} \subseteq -\mathcal{C}$, it is not hard to show that $\bar{f}(x) \geq f'(x)$. For the reverse inequality let $\varepsilon > 0$. By definition of f' we may find $g \in G_f$ such that $g(x) < f'(x) + \varepsilon/2$. Furthermore, as g is concave and continuous, Proposition 3.5 tells us that $g = \bar{g}$. Hence we may find $h \in \mathcal{A}$, $h \geq g$ such that

$$h(x) < \bar{g}(x) + \varepsilon/2 = g(x) + \varepsilon/2.$$

Combining the two estimates and using that $h \geq g > f$ yields

$$\bar{f}(x) \leq h(x) < g(x) + \varepsilon/2 < f'(x) + \varepsilon.$$

From this we conclude $\bar{f} = f'$ as desired.

G_f is directed downwards as $g_1, g_2 \in G_f$ implies $\min(g_1, g_2) \in G_f$, which is an immediate consequence of Lemma 5.3. The last statement follows from the preceding lemma: $\bar{f}$ is L^1 since it is bounded and measurable, and we obviously have $G_f \subseteq C(X)$. $\square$

This rather technical lemma permits a short proof of the converse to Proposition 4.8, originally due to Mokobodzki:

Proposition 6.3 (Mokobodzki)**.** *Let $\mu \in \mathcal{M}(X)$. Then μ is maximal if and only if $\mu(f) = \mu(\bar{f})$ for all $f \in \mathcal{C}$ if and only if $\mu(f) = \mu(\bar{f})$ for all $f \in C(X)$.*

Proof. Proposition 4.8 states that μ being maximal implies the two other statements, so we prove the converse. So let $\mu(f) = \mu(\bar{f})$ for all $f \in \mathcal{C}$. To show that μ is maximal, let $\lambda \in \mathcal{M}(X)$ be maximal with $\lambda \succ \mu$, the existence of which is ensured by Proposition 4.7.

For $g \in -\mathcal{C}$ we of course have $-g \in \mathcal{C}$, which implies $\lambda(g) \leq \mu(g)$ since $\lambda \succ \mu$. For any $f \in \mathcal{C}$ we then have

$$\lambda(f) = \lambda(\bar{f}) = \inf\{\lambda(g) \mid g \in G_f\} \leq \inf\{\mu(g) \mid g \in G_f\} = \mu(\bar{f}) = \mu(f).$$

Here we used Proposition 4.8 for the first equality, the preceding lemma for the second and second to last and the hypothesis for the last one. This is exactly the definition of $\lambda \prec \mu$, from which we by maximality obtain $\lambda = \mu$. Here we assumed that λ is maximal, but if ν is any measure with $\nu \succ \mu$, we may find a maximal measure $\lambda \succ \nu$. By transitivity $\lambda \succ \mu$ and by what has been show $\mu = \lambda$, hence $\nu = \mu$ as desired. $\square$

6.2 Towards uniqueness

The next lemma is a modification of Proposition 4.4. We recall that a measure $\mu \in \mathcal{M}^1(X)$ is *discrete* if $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$ for $x_i \in X$ and $\alpha_i \geq 0$ with $\sum_{i=1}^n \alpha_i = 1$.

Lemma 6.4. *Let $f \in C(X)$ and $x \in X$. Then $\bar{f}(x) = \sup\{\mu(f) \mid \mu \sim \delta_x, \mu \text{ discrete}\}$.*

Proof. Let $f \in C(X)$, $x \in X$ and $\varepsilon > 0$. In view of Proposition 4.4 it suffices to show that for each $\mu \sim \delta_x$ there exists a discrete probability measure $\lambda \sim \mu \sim \delta_x$ such that $\lambda(f) \geq \mu(f) - \varepsilon$. The idea of the proof is to partition X into finitely many regions using compactness and then replace the restrictions of μ to these regions with the Dirac measures of their resultants.

For any $y \in X$ we have $y \in U_y := f^{-1}(B(f(y), \varepsilon/4))$, which is open by continuity of f. For any $y_1, y_2 \in U_y$ we then have

$$|f(y_1) - f(y_2)| \leq |f(y_1) - f(y)| + |f(y) - f(y_2)| < \varepsilon/4 + \varepsilon/4 = \varepsilon/2.$$

By Lemma 1.6 we may then find a compact convex neighbourhood $C'_y \subseteq U_y$ of y. Using compactness of X we find finitely many C'_{y_i} covering X and define $C_1 = C'_{y_1}$ and

$$C_i = C'_{y_i} \setminus \bigcup_{j=1}^{i-1} C_i.$$

We now assume all C_i have positive measure, since if this is not the case one may simply discard those of measure 0 and renumber. As all C_i are Borel we may for any $B \in \mathbb{B}(X)$ define

$$\mu_i(B) = \mu(C_i)^{-1}\mu(C_i \cap B).$$

These are probability measures on C_i, which are subsets of the compact convex sets C'_{y_i}. Hence $r(\mu_i) = x_i \in C'_i$ by Proposition 2.7. We then define the discrete measure

$$\lambda = \sum_{i=1}^{n} \mu(C_i)\delta_{x_i}.$$

To conclude the proof we need to show $\lambda \sim \delta_x$ and $\mu(f) - \lambda(f) < \varepsilon$. The first claim is painless because of Corollary 4.3: It suffices to consider $g \in E^*_{|X}$. For these we have

$$\lambda(g) = \sum_{i=1}^{n} \mu(C_i)\delta_{x_i}(g) = \sum_{i=1}^{n} \mu(C_i)\mu_i(g) = \sum_{i=1}^{n} \int_{C_i} g\mathrm{d}\mu = \mu(g).$$

Here we used that $r(\delta_{x_i}) = r(\mu_i)$ for the first equality, the fact that $\mu_i = \mu(C_i)^{-1}\mu_{|C_i}$ for the second and that the C_i partition X (except possibly a set of measure 0) for the last. Hence $\lambda \sim \mu$, and $\lambda \sim \delta_x$ by transitivity.

The second claim requires a slightly longer, but also straightforward computation. Keeping in mind that $C_i \subseteq C'_{y_i}$ and $|f(y_1) - f(y_2)| < \varepsilon/2$ for all $y_1, y_2 \in C'_{y_i}$, we see that

$$\begin{aligned}
\mu(f) - \lambda(f) &= \sum_{i=1}^{n} \mu(C_i) \int_{C_i} f\mathrm{d}\mu_i - \sum_{i=1}^{n} \mu(C_i)\delta_{x_i}(f) = \sum_{i=1}^{n} \mu(C_i) \left(\int_{C_i} f\mathrm{d}\mu_i - f(x_i) \right) \\
&= \sum_{i=1}^{n} \mu(C_i) \left(\int_{C_i} f(y) - f(x_i)\mathrm{d}\mu_i(y) \right) \leq \sum_{i=1}^{n} \mu(C_i) \int_{C_i} \varepsilon/2\mathrm{d}\mu_i \\
&= \frac{\varepsilon}{2} \sum_{i=1}^{n} \mu(C_i)\mu_i(C_i) = \frac{\varepsilon}{2}.
\end{aligned}$$

This finishes the proof. □

We saw in Corollary 4.3 that if μ and λ coincide on $E^*_{|X}$, they coincide on all continuous affine functions. As a very last stepping stone towards the uniqueness theorem, we prove that if one of them is a Dirac measure, the assumption of continuity can be weakened to upper semicontinuity:

Lemma 6.5. *Let $f : X \to \mathbb{R}$ be affine and upper semicontinuous and let $x \in X$. If $\mu \in \mathcal{M}^1(X)$ and $\mu \sim \delta_x$, then $\mu(f) = f(x)$.*

Proof. Let $\mu \in \mathcal{M}^1(X)$ with $\mu \sim \delta_x$. We want to invoke Lemma 6.1 with $G = \{h \in \mathcal{A} \mid h > f\}$, as this implies

$$\mu(f) = \inf\{\mu(h) \mid h \in G\} = \inf\{h(x) \mid h \in G\} = f(x),$$

where the first equality is provided by the lemma and the inner stems from $\mu \sim \delta_x$. Hence we argue that the assumptions of the lemma hold.

As a first step we note that $f \in L^1$. Indeed, since f is upper semicontinuous the preimage $f^{-1}((-\infty, n))$ is open in X for any $n \in \mathbb{N}$, and compactness of X supplies a finite subcover, implying boundedness of f. We also remind ourselves that upper semicontinuity implies measurability.

Elements of G are of course continuous, so we need to argue that $f(x) = \inf\{h(x) \mid h \in G\}$ and that G is directed downwards. We start with the latter statement. Letting $h_1, h_2 \in G$, we denote their graphs in $E \times \mathbb{R}$ by J_1 and J_2, respectively. Since h_1 and h_2 are continuous and affine, J_1 and J_2 are compact and convex. Using convexity of J_i, it is not difficult to see that $\operatorname{co}(J_1 \cup J_2)$ is the image of the continuous map $\Phi : J_1 \times J_2 \times [0, 1] \to E \times \mathbb{R}$ defined by

$$(x, y, \alpha) \mapsto \alpha x + (1 - \alpha)y,$$

which implies compactness of $\operatorname{co}(J_1 \cup J_2)$. Furthermore, $\operatorname{co}(J_1 \cup J_2)$ is disjoint from the hypograph H_f of f, since for any $\alpha_i \geq 0$ with $\sum_{i=1}^n \alpha_i = 1$ we have

$$\sum_{i=1}^n \alpha_i h_i(x_i) > \sum_{i=1}^n \alpha_i f(x_i) = f\left(\sum_{i=1}^n \alpha_i x_i\right),$$

using f being affine. This means that the element $\sum_{i=1}^n \alpha_i(x_i, h_i(x_i)) \in \operatorname{co}(J_1 \cup J_2)$ lies strictly above the graph of f.

Since f is upper semicontinuous and concave, its hypograph is closed and convex. We may therefore use the Hahn-Banach separation theorem to separate H_f and and $\operatorname{co}(J_1 \cup J_2)$, i.e., find $L \in (E \times \mathbb{R})^*$ and $\lambda \in \mathbb{R}$ such that

$$\sup L(H_f) < \lambda < \inf L(\operatorname{co}(J_1 \cup J_2)).$$

Following the line of reasoning in Lemma 3.4, we may now define a new continuous affine function h_3 on X by the equation $L(x, h_3(x)) = \lambda$. By construction we will then have $h_3 > f$ and $h_3 < h_i$ for $i \in \{1, 2\}$. Hence G is in fact directed downwards.

The proof that $f(x) = \inf\{h(x) \mid h \in G\}$ is similar to the one above and in fact almost identical to the proof of Lemma 3.4: Defining f' to be the right hand side we have $f \leq f'$ by construction. Assuming $f(x_0) < f'(x_0)$ for some x_0, we may separate H_f and the point $(x_0, f'(x_0)) \in E \times \mathbb{R}$ with a continuous linear functional L; the resulting affine function h will satisfy both $h \in G$ and $h(x_0) < f'(x_0)$, a contradiction.

We have now argued that Lemma 6.1 does indeed apply, finishing the proof. ☐

6.3 The Choquet-Meyer theorem

After this large amount of preparatory work we have finally paved the way for proving the desired uniqueness result:

Theorem 6.6 (Choquet-Meyer)**.** *Let X be a non-empty, compact and convex subset of a Hausdorff LCS, $\mathcal{Q}^1$ the set of maximal probability measures on X and $\mathcal{C}$ the set of continuous convex functions on X. Then the following are equivalent:*

1. *X is a Choquet simplex*
2. *For each $f \in \mathcal{C}$, $\bar{f}$ is affine on X*
3. *If $\mu \in \mathcal{Q}^1$ with $x = r(\mu)$ and $f \in \mathcal{C}$, then $\bar{f}(x) = \mu(f)$*
4. *For all $f, g \in \mathcal{C}$, $\overline{f + g} = \bar{f} + \bar{g}$.*

5. *For each $x \in X$ there is a unique $\mu_x \in \mathcal{Q}^1$ such that $r(\mu_x) = x$*

Proof. 1. $\implies$ *2.* Assume X is a Choquet simplex and let $f \in \mathcal{C}$. We know that $\bar{f}$ is concave, so we only need to show that it is also convex. Let $x_1, x_2 \in X$, $\alpha_1 \in (0,1)$ and $\alpha_2 = 1 - \alpha_1$. Defining $z = \alpha_1 x_1 + \alpha_2 x_2$ we want to show

$$\bar{f}(z) = \bar{f}(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 \bar{f}(x_1) + \alpha_2 \bar{f}(x_2).$$

For this step we will use two lemmas we have established earlier: The decomposition lemma for vector lattices and Lemma 6.4. Due to the second lemma it suffices to show that the right hand side of the above inequality is an upper bound for for $\{\mu(f) \mid \mu \sim \delta_z,\ \mu \text{ discrete}\}$.

Now let $\mu \in \mathcal{M}^1(X)$ be discrete with $\mu \sim \delta_z$. Then $\mu = \sum_{j=1}^n \beta_j \delta_{y_j}$ some $y_j \in X$ and $\beta_j \geq 0$ with $\sum_{j=1}^n \beta_j = 1$. We assume without loss of generality that $\beta_j > 0$ for all j. Since the resultant map is affine, we get

$$r(\mu) = z = \sum_{j=1}^n \beta_j r(\delta_{y_j}) = \sum_{j=1}^n \beta_j y_j.$$

This means we have

$$\alpha_1 x_1 + \alpha_2 x_2 = z = \sum_{j=1}^n \beta_j y_j.$$

Since X is a Choquet simplex, $\tilde{X} - \tilde{X}$ is a vector lattice in the ordering induced by $\tilde{X}$, so the decomposition lemma holds. Furthermore, for any $w \in \tilde{X} - \tilde{X}$ we have $w \geq 0$ if and only if $w \in \tilde{X}$. This implies $\alpha_i x_i \geq 0$ and $\beta_j y_j \geq 0$, so from the lemma we obtain $z'_{ij} \in \tilde{X}$ such that $\alpha_i x_i = \sum_{j=1}^n z'_{ij}$ and $\beta_j y_j = z'_{1j} + z'_{2j}$. Using the definition of $\tilde{X}$ we may write $z'_{ij} = \gamma_{ij} z_{ij}$ for $\gamma_{ij} \geq 0$ and $z_{ij} \in X$. We then have

$$x_i = \alpha_i^{-1} \sum_{j=1}^n \gamma_{ij} z_{ij} = \sum_{j=1}^n \alpha_i^{-1} \gamma_{ij} z_{ij}.$$

Since $X \subseteq L^{-1}(1)$ for some $L \in E^*$, we see that

$$1 = L(x_i) = \sum_{j=1}^n \alpha_i^{-1} \gamma_{ij} L(z_{ij}) = \sum_{j=1}^n \alpha_i^{-1} \gamma_{ij},$$

implying that $\sum_{j=1}^n \alpha_i^{-1} \gamma_{ij} = 1$. This leads us to defining the discrete measures

$$\mu_i = \sum_{j=1}^n \alpha_i^{-1} \gamma_{ij} \delta_{z_{ij}},$$

which satisfy $r(\mu_i) = x_i$. By Corollary 4.3 this means $\mu_i \sim \delta_{x_i}$, and once again using Lemma 6.4 we see that

$$\bar{f}(x_i) \geq \mu_i(f) = \sum_{j=1}^n \alpha_i^{-1} \gamma_{ij} f(z_{ij}).$$

On the other hand we have $y_j = \beta_j^{-1} \gamma_{1j} z_{1j} + \beta_j^{-1} \gamma_{2j} z_{2j}$, where we similarly have $\beta_j^{-1} \gamma_{1j} + \beta_j^{-1} \gamma_{2j} = 1$. Convexity of f then implies

$$f(y_j) = f(\beta_j^{-1} \gamma_{1j} z_{1j} + \beta_j^{-1} \gamma_{2j} z_{2j}) \leq \beta_j^{-1} \gamma_{1j} f(z_{1j}) + \beta_j^{-1} \gamma_{2j} f(z_{2j}).$$

Combining the estimates we finally see that

$$\begin{aligned}
\mu(f) = \sum_{j=1}^n \beta_j \delta_{y_j}(f) &= \sum_{j=1}^n \beta_j f(y_j) \leq \sum_{j=1}^n \gamma_{1j} f(z_{1j}) + \gamma_{2j} f(z_{2j}) \\
&= \alpha_1 \sum_{j=1}^n \alpha_1^{-1} \gamma_{1j} f(z_{1j}) + \alpha_2 \sum_{j=1}^n \alpha_2^{-1} \gamma_{2j} f(z_{2j}) = \alpha_1 \mu_1(f) + \alpha_2 \mu_2(f) \\
&\leq \alpha_1 \bar{f}(x_1) + \alpha_2 \bar{f}(x_2).
\end{aligned}$$

Since μ was an arbitrary discrete measure with $\mu \sim \delta_z$, we see that

$$\bar{f}(\alpha_1 x_1 + \alpha_2 x_2) = \bar{f}(z) = \sup\{\mu(f) \mid \mu \sim \delta_z,\ \mu \text{ discrete}\} \leq \alpha_1 \bar{f}(x_1) + \alpha_2 \bar{f}(x_2),$$

so $\bar{f}$ is convex and hence affine.

2. $\implies$ *3.* Let $\mu \in \mathcal{Q}^1$ and $f \in \mathcal{C}$. Then $\bar{f}$ is upper semicontinuous by Proposition 3.5 and affine by hypothesis. Since $r(\mu) = x$ is equivalent to $\mu \sim \delta_x$, we see by Lemma 6.5 that

$$\bar{f}(x) = \mu(\bar{f}) = \mu(f),$$

where we use maximality of μ for the last equality.

3. $\implies$ *4.* Let $f, g \in \mathcal{C}$ and let $x \in X$. By Lemma 4.7 there exists a maximal measure μ such that $\mu \succ \delta_x$, which in particular means $\mu \sim \delta_x$ and hence $r(\mu) = x$. Using the hypothesis twice we see that

$$\overline{(f+g)}(x) = \mu(f+g) = \mu(f) + \mu(g) = \bar{f}(x) + \bar{g}(x).$$

4. $\implies$ *5.* Let $x_0 \in X$. Defining $T' : \mathcal{C} \to \mathbb{R}$ by $f \mapsto \bar{f}(x_0)$, we know from Proposition 3.5 that T' is positive homogenous, and it is additive by hypothesis. We may therefore extend T' to $T : \mathcal{C} - \mathcal{C} \to \mathbb{R}$ by $T(f - g) = T'(f) - T'(g)$. It is both straightforward and uninteresting to verify that T' being additive and positive homogenous ensures T being linear.

Using property 3. of Proposition 3.5, we see immediately that

$$|T(f-g)| = |\bar{f}(x) - \bar{g}(x)| \leq \|f - g\|_\infty,$$

from which we deduce that T is bounded with $\|T\|_\infty \leq 1$. As we have used several times before, the subspace $\mathcal{C} - \mathcal{C}$ is dense in $C(X)$ by the lattice version of the Stone-Weierstrass theorem, so we may extend T by continuity to a linear functional $L \in C(X)^*$ with $\|L\|_\infty \leq 1$.

As per usual when we produce a measure, we want to invoke Riesz' representation theorem. It is obvious that $L(1) = T(1) = \bar{1}(x_0) = 1$, so the measure would in fact be a probability measure. We still need to argue that L is positive. We note that if $\mathcal{C} - \mathcal{C} \ni g_1 - g_2 \geq 0$, we also have $L(g_1 - g_2) \geq 0$ since the upper envelope is monotone.

Now for an arbitrary $f \in C(X)$ with $f \geq 0$ it is not a priori obvious that f may be approximated by non-negative elements of $\mathcal{C} - \mathcal{C}$. But for any $\varepsilon > 0$ we may approximate f with a strictly positive function, say $f + \frac{1}{N}$ for sufficiently large $N \in \mathbb{N}$. By the definition of the $\infty-$norm and since $\mathcal{C} - \mathcal{C}$ is dense, we may find strictly positive $g_1 - g_2 \in \mathcal{C} - \mathcal{C}$ with $\|f + 1/N - (g_1 - g_2)\|_\infty < \varepsilon$, which by the triangle inequality is sufficiently close to f. Thus f may indeed be approximated by positive $g_1 - g_2$ for which $L(g_1 - g_2) \geq 0$, and by continuity we infer $L(f) \geq 0$.

Hence L is in fact a positive linear functional on $C(X)$, for which we find $\nu \in \mathcal{M}^1(X)$ such that $L(f) = \int_X f \mathrm{d}\nu$ for all $f \in C(X)$. Then, for each $f \in \mathcal{C}$, we have

$$\nu(f) = L(f) = \bar{f}(x_0).$$

By Proposition 4.4 we see that

$$\nu(f) = \sup\{\mu(f) \mid \mu \sim \delta_{x_0}\}.$$

Since $\mu \succ \delta_{x_0}$ if and only if $\mu \sim \delta_{x_0}$, any maximal λ with $\lambda \sim \delta_{x_0}$ satisfies $\nu(f) \geq \lambda(f)$. As this holds for any $f \in \mathcal{C}$, we see that $\nu \succ \lambda$, which by maximality of λ implies $\nu = \lambda$. Hence ν is the unique maximal probability measure with $\nu \sim \delta_{x_0}$, i.e., $r(\nu) = x_0$.

5. $\implies$ *1.* We know from Lemma 2.6 that the resultant map $r : \mathcal{M}^1(X) \to X$ is affine, and therefore its restriction to $\mathcal{Q}^1$ also is. Furthermore, the existence of a maximal $\mu \succ \delta_x$ implies that it is surjective, and the hypothesis now yields that it is also injective. Hence we have an affine bijection between $\mathcal{Q}^1$ and X, and from Lemma 5.14 we know that the former is a Choquet simplex. By Lemma 5.9 this affine bijection will extend to an additive, order preserving bijection between $\tilde{\mathcal{Q}}^1$ and $\tilde{X}$, carrying with it the property of being a lattice. Hence X is in fact a Choquet simplex, and we conclude the proof. □

Since maximal measures are characterised by the property that $\mu(f) = \mu(\overline{f})$ for each $f \in C(X)$, they are supported by the sets $\{x \in X \mid f(x) = \overline{f}(x)\}$. The following proposition illuminates how this is intimately connected to the extreme points of X:

Proposition 6.7. *Let X be compact convex. Then $Ext(X) = \bigcap_{f \in C(X)} \{x \in X \mid f(x) = \overline{f}(x)\}$.*

Proof. Let $x \in \mathrm{Ext}(X)$. Then, by Proposition 2.8, the only measure representing x is δ_x. Furthermore, for any $f \in C(X)$ we have

$$\overline{f}(x) = \sup\{\mu(f) \mid \mu \sim \delta_x\}$$

by Proposition 4.4. Since μ representing x and $\mu \sim \delta_x$ are equivalent, we see $f(x) = \overline{f}(x)$.

Conversely, assume $f(x) = \overline{f}(x)$ for every $f \in C(X)$ and $x = \alpha y + (1-\alpha)z$ for some $\alpha \in (0,1)$ and $y, z \in X$. Then, for each $f \in \mathcal{C}$, we have

$$f(x) \leq \alpha f(y) + (1-\alpha) f(z) \leq \alpha \overline{f}(y) + (1-\alpha)\overline{f}(z) \leq \overline{f}(x),$$

using concavity of $\overline{f}$ for the last inequality. Since $f(x) = \overline{f}(x)$, we must have equality in all of the above. In particular this entails $f(x) = \alpha f(y) + (1-\alpha) f(z)$ for every $f \in \mathcal{C}$. Multiplying the equation by -1 it also holds for every $f \in -\mathcal{C}$, and by linearity this extends to all $g \in \mathcal{C} - \mathcal{C}$. Since this set is uniformly dense in $C(X)$ and the vector space operations are continuous, this identity actually holds for every $f \in C(X)$.

With this identity in hand we can easily see that x must be extreme in X. Indeed, if $x \neq y$, we can apply Urysohn's lemma to find a continuous $f : X \to [0,1]$ with $f(x) = 1$ and $f(y) = 0$. Using the identity we see that

$$1 = (1-\alpha) f(z),$$

which is impossible since the right hand side is strictly smaller than 1. □

We are now ready to prove an interesting consequence of the Choquet-Meyer theorem, which is Choquet's original uniqueness theorem for metrizable X. We begin with the following corollary:

Corollary 6.8. *Let X be a Choquet simplex such that $Ext(X)$ is Baire. Then for each $x \in X$ there exists a unique probability measure μ which represents x and is supported by $Ext(X)$.*

Proof. Let $x_0 \in X$. By (the proof of) the Choquet-Bishop-De Leeuw theorem, there exists a maximal probability measure μ representing x_0 which is supported by every Baire set containing $\mathrm{Ext}(X)$. Since $\mathrm{Ext}(X)$ is assumed Baire, this means in particular that μ is supported by $\mathrm{Ext}(X)$.

For uniqueness assume λ also represents x_0 and $\lambda(\mathrm{Ext}(X)) = 1$. By the above proposition we have $\mathrm{Ext}(X) \subseteq \{x \in X \mid f(x) = \overline{f}(x)\}$ for each $f \in C(X)$, implying that

$$\lambda(\{x \in X \mid f(x) \neq \overline{f}(x)\}) = 0.$$

From this we may infer $\lambda(f) = \lambda(\overline{f})$, which by Proposition 6.3 is equivalent to λ being maximal. Then by the Choquet-Meyer theorem we have $\lambda = \mu$. □

Remark. It is worth noting that we also proved that if $\mu(\mathrm{Ext}(X)) = 1$ for a probability measure μ, then μ is maximal.

Choquet's uniqueness theorem is then simply a special case of the above:

Theorem 6.9 (Choquet)**.** *Let X be compact, convex and metrizable. Then X is a Choquet simplex if and only if each $x \in X$ is represented by a unique probability measure supported by the extreme points of X.*

Proof. Assume X is a Choquet simplex. Then the result follows from the above corollary, since $\mathrm{Ext}(X)$ is Baire by Proposition 3.8. Conversely, assume for any $x_0 \in X$ there exists a unique $\mu \in \mathcal{M}^1(X)$ with $\mu(\mathrm{Ext}(X)) = 1$ which represents x_0. Let λ be a maximal probability measure with $\lambda \sim \delta_{x_0}$. Then, once again appealing to the proof of the Choquet-Bishop-De Leeuw theorem, $\lambda(\mathrm{Ext}(X)) = 1$. Thus $\lambda = \mu$, and we infer that X is a Choquet simplex by the Choquet-Meyer theorem. □

6.4 Different notions of simplices

In the following example we describe an important class of Choquet simplices:

Example 6.10 (Bauer simplices)**.** A Choquet simplex X is a *Bauer simplex* if $\mathrm{Ext}(X)$ is closed. We have already seen in Proposition 5.13 that if K is compact Hausdorff, $\mathcal{M}^1(K)$ is a Choquet simplex. In Proposition 2.12 we proved that $\mathrm{Ext}(\mathcal{M}^1(K)) = \{\delta_x \mid x \in K\}$, and furthermore that the embedding $x \mapsto \delta_x$ is a homeomorphism. Therefore $\mathcal{M}^1(K)$ is always a Bauer simplex.

Even more interestingly, the above example is in a certain sense universal. Indeed, if X is a compact Bauer simplex, $\mathrm{Ext}(X)$ is closed and hence compact. Furthermore, the resultant map $r : \mathcal{M}^1(\mathrm{Ext}(X)) \to X$ is affine and continuous. By the reformulated Krein-Milman theorem, we know that there exists a representing measure supported by $\mathrm{Ext}(X)$, so r is also surjective. Then, by the remark following Corollary 6.8 and the Choquet-Meyer theorem, the representing measure is also unique. Thus r is furthermore injective, so by the usual compact-Hausdorff argument, r is an affine homeomorphism of $\mathcal{M}^1(\mathrm{Ext}(X))$ onto X. In particular, X may be identified with $\mathcal{M}^1(K)$ for $K = \mathrm{Ext}(X)$.

There exists a plethora of theoretical applications in other areas of analysis of compact convex sets with closed extreme points, some of which are Bauer simplices. Examples of applications include representations of positive harmonic functions, completely monotonic functions and positive definite functions on abelian groups. We refer the reader to chapter 14 in [6] for a survey.

We also want to supply the reader with another very intuitive characterisation of Choquet simplices. For this we note that $a + \alpha X$ for $a \in E$ and $\alpha > 0$ is a *homothetic image* of X. We then have the following:

Theorem 6.11. *Let X be compact and convex. Then X is a Choquet simplex if and only if the intersection of any two homothetic images of X is either empty, a point or once again a homothetic image of X.*

This characterisation is probably the one that is most easily visualised. It also emphasizes the relation to self-similarity even more than the characterisation using intersections of cones discussed in section 5. We refer the reader to Theorem 3.2 in [4] for one direction of the proof and further references.

We conclude this section by proving that this new notion of a Choquet simplex coincides with the usual notion in finite dimensional spaces. A common definition of an n-dimensional simplex is that X is the convex hull of $n+1$ affinely independent points. We will need the following theorem attributed to Minkowski, a proof of which can be found on p. 8 in [6]:

Theorem 6.12. *Let X be a compact convex subset of a finite-dimensional LCS E. Then $X = co(Ext(X))$.*

The proof referenced above technically only applies to subsets of $\mathbb{R}^n$, but by Theorem 3.7 in [11], any finite-dimensional topological vector space is linearly homeomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N}$. Thus X is affinely homeomorphic to a compact convex subset of $\mathbb{R}^n$. We turn to the mentioned theorem:

Theorem 6.13. *Let X be compact convex such that $E = \tilde{X} - \tilde{X}$ has dimension n. Then X is the convex hull of n affinely independent points if and only if X is a Choquet simplex.*

Proof. Throughout the proof we assume that $n \geq 2$, since $n = 1$ implies that X is just a point, for which the result is trivial. Assume first that X is the convex hull of n affinely independent points $x_1, \dots, x_n$. Since X is contained in a closed hyperplane missing the origin, say $L^{-1}(1)$ for a suitable $L \in E^*$, the points are in fact also linearly independent. Indeed, if $\sum_{i=1}^n \lambda_i x_i = 0$, then $\sum_{i=1}^n \lambda_i = 0$ by applying L to both sides of the equation. Thus $\lambda_1 = -\sum_{i=2}^n \lambda_i$, from which we deduce

$$0 = \sum_{i=1}^n \lambda_i x_i = \sum_{i=2}^n \lambda_i x_i - \sum_{i=2}^n \lambda_i x_1 = \sum_{i=2}^n \lambda_i (x_i - x_1).$$

Affine independence then yields $\lambda_i = 0$ for all $i \in \{2, \ldots, n\}$, which also implies $\lambda_1 = 0$. Thus $x_1, \ldots, x_n$ are indeed also linearly independent.

Since $x_1, \ldots, x_n$ are linearly independent and span $\tilde{X} - \tilde{X} = E$, they form a basis. We can then define $T : E \to \mathbb{R}^n$ by $x_i \mapsto e_i$, where $e_1, \ldots e_n$ is the canonical basis of $\mathbb{R}^n$. Since T is a linear isomorphism, it is in particular an affine bijection from X onto its image. Furthermore, by construction of T, $\widetilde{T(X)}$ consists of the elements of $\mathbb{R}^n$ which have non-negative coordinates. This is obviously a convex cone with base $T(X)$, and if $y, y' \in \widetilde{T(X)}$, it is easy to verify that taking the maximum of each coordinate yields a least upper bound in the ordering induced by $\widetilde{T(X)}$. Hence $T(X)$ is a Choquet simplex, and since T is an affine bijection, we appeal to Proposition 5.9 to conclude that X is a Choquet simplex as well.

Conversely, assume that X is a Choquet simplex. By Minkowski's theorem, X is the convex hull of its extreme points. Then the extreme points span E, so we can find $x_1, \ldots, x_n \in \mathrm{Ext}(X)$ which are a basis for E. Since being linearly independent is stronger than being affinely independent, X will be the convex hull of n affinely independent points if we can show that there are no more extreme points.

To this end assume there is one more extreme point x_{n+1}. Then all extreme points cannot be linearly independent, i.e, there exist coefficients $\lambda_i \in \mathbb{R}$ such that $\sum_{i=1}^{n+1} \lambda_i x_i = 0$ and not all coefficients are 0. By multiplying with -1 if need be, we may assume there exists a strictly positive coefficient. We then partition the indices into those having negative and non-negative coefficients, that is, let

$$P = \{i \in \{1, \ldots n+1\} \mid \lambda_i \geq 0\} \quad \text{and} \quad N = \{i \in \{1, \ldots n+1\} \mid \lambda_i < 0\}.$$

Then $\lambda := \sum_{i \in P} \lambda_i > 0$ and since $X \subseteq L^{-1}(1)$, we have

$$L\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) = \sum_{i=1}^{n+1} \lambda_i = 0,$$

which implies $\sum_{i \in N} \lambda_i = -\lambda$. We then define $x = \sum_{i \in P} \lambda^{-1} \lambda_i x_i$. This is a convex combination of extreme points, as

$$1 = L(x) = \sum_{i \in P} \lambda^{-1} \lambda_i.$$

Since $\sum_{i \in P} \lambda_i x_i + \sum_{i \in N} \lambda_i x_i = 0$, we have

$$x = \lambda^{-1} \left(\sum_{i \in N} -\lambda_i x_i\right).$$

This is also a convex combination as $\lambda = -\sum_{i \in N} \lambda_i$ and $-\lambda_i \geq 0$ when $i \in N$. Since all x_i are distinct, the two discrete measures $\sum_{i \in P} \lambda^{-1} \lambda_i \delta_{x_i}$ and $\sum_{i \in N} \lambda^{-1}(-\lambda_i) \delta_{x_i}$ are distinct and represent x. As they both measure a subset of $\mathrm{Ext}(X)$ to 1, we may by the exact same argument used for uniqueness in Corollary 6.8 deduce that they must be maximal. But this violates the Choquet-Meyer theorem. $\square$

7 Applications to ergodic theory

In this section we will apply the Choquet-Bishop-DeLeeuw and Choquet-Meyer theorems to the study of ergodic measures. Compared to many other applications of Choquet theory, this one is particularly interesting to us, since the set in question need not be metrizable and the extreme points need not be closed. Therefore we need the full strength provided by the generality of the two aforementioned theorems. Most results of this section can be found in chapter 12 of [8], although we have preferred the use of Radon measures over that of Baire measures. Letting $T(\mu)$ denote the image measure of μ under a measurable map T, we begin with a few definitions that are central in ergodic theory:

Definition 7.1. Let $(K, \mathbb{A})$ be a measureable space and let $(T_i)_{i\in I} : K \to K$ be a family of measurable maps. A measure μ on K is *invariant* if $\mu = T_i(\mu)$ for all $i \in I$.

Invariant measures preserve, unsurprisingly, the measure of a set when taking preimages under T_i. It is also useful to formalize the notion of the set itself being preserved, not just its size:

Definition 7.2. Let μ be a measure on $(K, \mathbb{A})$. A set $A \in \mathbb{A}$ is *invariant* $\pmod{\mu}$ if $\mu(A \triangle T_i^{-1} A) = 0$ for all $i \in I$. Furthermore, let $\mathbb{A}_\mu$ denote the set of sets which are invariant $\pmod{\mu}$.

Remark. Using the two set identities

$$A \triangle B = A^c \triangle B^c \quad \text{and} \quad \left(\bigcup_{n=1}^{\infty} A_n \right) \triangle \left(\bigcup_{n=1}^{\infty} B_n \right) \subseteq \bigcup_{n=1}^{\infty} A_n \triangle B_n$$

it is straightforward to prove that $\mathbb{A}_\mu$ is a σ-algebra.

Next we define the notion of ergodic measures, a special type of invariant measures. There are several equivalent definitions in the literature, and we will be working with the formulation involving $\mathbb{A}_\mu$:

Definition 7.3. Let $(K, \mathbb{A})$ be a measureable space and let $(T_i)_{i\in I} : K \to K$ be a family of measurable maps. An invariant measure μ is *ergodic* if $\mu(A) \in \{0, 1\}$ for all $A \in \mathbb{A}_\mu$.

In the following we will see that the ergodic measures are exactly the extreme points in the set of invariant measures. Before we focus on a specific topological space, we need a couple of measure theoretic lemmas. The first lemma relates the property of being invariant to a property of the Radon-Nikodym derivative. While the proof is quite elegant, we omit it and refer the reader to Lemma 12.1 in [8].

Lemma 7.4. *Let μ and ν be measures on $(K, \mathbb{A})$ such that ν is invariant and μ is absolutely coontinuous with respect to ν. Futhermore, let f denote the Radon-Nikodym derivative of μ with respect to ν. Then μ is invariant if and only if $f = f \circ T_i$ ν-a.e. for all $i \in I$.*

The next lemma will help us prove that the ergodic measures are extreme points:

Lemma 7.5. *Let $(K, \mathbb{A})$ be a measureable space and let $(T_i)_{i\in I} : K \to K$ be a family of measurable maps. Let μ and ν be invariant finite measures on K which coincide on $\mathbb{A}_{\mu+\nu}$. Then $\mu = \nu$.*

Proof. Let μ and ν have Radon-Nikodym derivatives f and g with respect to $\mu+\nu$. We prove that these are equal $(\mu + \nu)-$a.e, from which the conclusion follows. We note for later use that since μ and ν are finite, $f, g \in L^1(\mu + \nu)$.

To be able to use the hypothesis, we first claim that f and g are $\mathbb{A}_{\mu+\nu}$ measurable. Indeed, for $r \in \mathbb{R}$ let $I_r = (\infty, r)$, the collection of which generates $\mathbb{B}(\mathbb{R})$. Then

$$f^{-1}(I_r) \triangle T_i^{-1}(f^{-1}(I_r)) \subseteq \{x \in K \mid f(x) \neq f(T_i(x))\},$$

since if $x \in f^{-1}(I_r) \setminus T_i^{-1}(f^{-1}(I_r))$ then $f(x) < r$ while $f(T_i(x)) \geq r$ and vice versa. Since both μ and ν are invariant, $\mu + \nu$ is as well. Lemma 7.4 then implies that $f = f \circ T_i$ a.e., so

$f^{-1}(I_r) \triangle T^{-1}(f^{-1}(I_r)$ is contained in a $\mu+\nu$ null set. As this is true for all $i \in I$, we see that $f^{-1}(I_r) \in \mathbb{A}_{\mu+\nu}$. Interchanging f and g we see that they are both $\mathbb{A}_{\mu+\nu}$ measureable.

To finish the proof let $A = \{x \in K \mid f(x) > g(x)\}$. By the above we have $A \in \mathbb{A}_{\mu+\nu}$, so applying the hypothesis yields

$$\int_A f \mathrm{d}(\mu+\nu) = \mu(A) = \nu(A) = \int_A g \mathrm{d}(\mu+\nu),$$

from which we deduce that the non-negative function $1_A(f-g)$ integrates to 0. But then it must be 0 almost everywhere, i.e, $(\mu+\nu)(A) = 0$. Since the same line of reasoning applies to $B = \{x \in K \mid g(x) > f(x)\}$, we conclude that f equals g $(\mu+\nu)$-a.e. $\square$

We will now turn our attention to the more specific case where K is a compact Hausdorff topological space and the measures in question are Radon. Let $\mathcal{I}(K)$ denote the set of invariant Radon measures and $\mathcal{I}^1(K)$ the set of invariant probability measures. The next proposition shows that we are in the setting of Choquet theory:

Proposition 7.6. *Let K be compact Hausdorff and let $(T_i)_{i\in I} : K \to K$ be a family of continuous maps which commute with respect to composition. Then the set of invariant probability measures $\mathcal{I}^1(K)$ is non-empty, compact and convex in the LCS $E = C(K)^*$ equipped with the weak* topology.*

Proof. Since the property of being an invariant probability measure is preserved by convex combinations, $\mathcal{I}^1(K)$ is convex. For each $i \in I$ define $T_i^* : \mathcal{M}^1(K) \to \mathcal{M}^1(K)$ by $\mu \mapsto T_i(\mu)$. We argue that T_i^* is weak* continuous. In general, by the change of variables formula for image measures, we have

$$(T_i(\mu))(f) = \int_K f \mathrm{d}(T_i(\mu)) = \int_{T_i^{-1}(K)} f \circ T_i \mathrm{d}\mu = \mu(f \circ T_i)$$

for any $f \in C(K)$. Now let $\mu_\alpha \to \mu$ and $f \in C(K)$. Using continuity of T_i and the usual characterisation of weak* convergence,

$$(T_i^*\mu)(f) = (T_i(\mu))(f) = \mu(f \circ T_i) = \lim_\alpha \mu_\alpha(f \circ T_i) = \lim_\alpha (T_i(\mu_\alpha))(f) = \lim_\alpha (T_i^*\mu_\alpha)(f),$$

from which we deduce that T_i^* is indeed continuous. Since

$$\mathcal{I}^1(K) = \{\mu \in M^1(K) \mid \forall i \in I : T_i^*(\mu) = \mu\},$$

we see that the invariant probability measures are exactly the fixed points of $(T_i^*)_{i\in I}$. As the set of fixed points of a family of continuous maps is closed, we conlude that $\mathcal{I}^1(K)$ is closed in $\mathcal{M}^1(K)$ and thus compact. It is not difficult to show that each T_i^* is also affine, so by the Markov-Kakutani fixed point theorem (see Theorem V.20 on p. 152 in [10]), $\mathcal{I}^1(K) \neq \emptyset$. $\square$

In addition to being non-empty, compact and convex, $\mathcal{I}^1(K)$ is also a Choquet simplex:

Proposition 7.7. *Let K be compact Hausdorff and let $(T_i)_{i\in I} : K \to K$ be a family of continuous maps which commute with respect to composition. Then $\mathcal{I}^1(K)$ is a Choquet simplex.*

Proof. The proof mirrors the proof of Proposition 5.13. Of course $\mathcal{I}(K)$ is a convex cone with base $\mathcal{I}^1(K)$. We show that $\mathcal{I}(K)$ has the greatest lower bound property and appeal to Proposition 5.10. Let $\mu, \nu \in \mathcal{I}(K)$ and let f and g denote the Radon-Nikodym derivatives with respect to $\mu+\nu$. Defining $h = \min(f,g)$ and $\lambda = h \cdot (\mu+\nu)$, we claim that λ is the least upper bound of μ and ν. We already saw that λ is Radon, and we furthermore claim it is invariant. Using invariance of $\mu+\nu$ and applying Lemma 7.4, we see that $f = f \circ T_i$ and $g = g \circ T_i$, forcing $h = h \circ T_i$ for each $i \in I$. Applying the lemma once more we see that λ is invariant.

Since $f \geq h$, we have

$$\mu - \lambda = f \cdot (\mu+\nu) - h \cdot (\mu+\nu) = (f-h) \cdot (\mu+\nu),$$

that is, $\mu - \lambda$ has density with respect to $\mu+\nu$. By the previous argument and since $(f-h) \circ T_i = f \circ T_i - h \circ T_i$, we also have $\mu - \lambda \in \mathcal{I}(K)$. Therefore $\lambda \leq \mu$ in the ordering induced by $\mathcal{I}(K)$, and similarly $\lambda \leq \nu$.

The fact that it is also the greatest lower bound follows from the same argument. $\square$

We are now almost ready to apply the existence and uniqueness theorems established in earlier sections. The only thing remaining is characterising $\mathrm{Ext}(\mathcal{I}^1(K))$.

Proposition 7.8. *Let K be compact Hausdorff and let $(T_i)_{i\in I} : K \to K$ be a family of continuous maps which commute with respect to composition. Then*

$$\mathcal{E} := \{\mu \in \mathcal{I}^1(K) \mid \mu \text{ ergodic}\} = \mathrm{Ext}(\mathcal{I}^1(K))$$

Proof. First assume $\mu \notin \mathcal{E}$. Then there exists some $A \in \mathbb{A}_\mu$ such that $0 < \mu(A) < 1$. We decompose μ to probability measures μ_1 and μ_2 supported by A and A^c, respectively: Let $B \in \mathbb{B}(K)$ and define

$$\mu_1(B) = \mu(A)^{-1}\mu(B \cap A), \qquad \mu_2(B) = \mu(A^c)^{-1}(B \cap A^c).$$

We then see that $\mu = \mu(A)\mu_1 + (1 - \mu(A))\mu_2$ is a proper convex combination. Thus, if we argue that $\mu_i \in \mathcal{I}^1(K)$, we may infer that μ is not extreme.

Since we know that μ_1 is Radon, it suffices to show to it is also invariant. By invariance of μ we see that $\mu(B\cap A) = \mu(T^{-1}(B)\cap T^{-1}(A))$, so we just need to show that $\mu(T^{-1}(B)\cap T^{-1}(A)) = \mu(T^{-1}(B) \cap A)$. As intersections distribute over symmetric differences, we see that

$$\big(T^{-1}(B) \cap T^{-1}(A)\big) \triangle \big(T^{-1}(B) \cap A\big) = T^{-1}(B) \cap (T^{-1}(A) \triangle A).$$

Since $A \in \mathbb{A}_\mu$, this implies that the symmetric difference of the sets in question has measure 0. Now, it is a general fact that sets whose symmetric difference measures to zero have the same measure. Indeed, if we for sets C and D have $\mu(C\triangle D) = 0$, we obtain

$$\mu(C \cup D) = \mu(C\triangle D) + \mu(C \cap D) = \mu(C \cap D).$$

As both C and D are squeezed in between $C \cup D$ and $C \cap D$, we see that their measures must coincide. Thus μ_1 is invariant, and since A is invariant (mod μ) if and only if A^c is, the same argument shows μ_2 is invariant.

Now suppose μ is ergodic and $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$ for $\alpha \in (0, 1)$ and $\mu_i \in \mathcal{I}^1(K)$. We claim that μ and μ_i coincide on $\mathbb{A}_\mu$. Indeed, if $A \in \mathbb{A}_\mu$, $\mu(A) \in \{0, 1\}$. Since $\mathrm{Ext}([0, 1]) = \{0, 1\}$, we see that

$$\mu(A) = \alpha\mu_1(A) + (1 - \alpha)\mu_2(A)$$

forces $\mu(A) = \mu_1(A) = \mu_2(A)$.

To finish the argument we note that if $(\mu + \mu_1)(A\triangle T^{-1}(A)) = 0$, then $\mu(A\triangle T^{-1}(A)) = 0$. Thus $\mathbb{A}_{\mu+\mu_1} \subseteq \mathbb{A}_\mu$, on which μ and μ_1 coincide. By Lemma 7.5 we conclude $\mu = \mu_1$, and since the same line of reasoning applies to μ and μ_2, we see that $\mu \in \mathrm{Ext}(\mathcal{I}^1(K))$. □

We are now ready to apply the two main theorems of the thesis:

Theorem 7.9. *Let K be compact Hausdorff and let $(T_i)_{i\in I} : K \to K$ be a family of commuting continuous maps. Then $\mathcal{I}^1(K) \neq \emptyset$, and for each $\mu \in \mathcal{I}^1(K)$ there exists a $m \in \mathcal{M}^1(\mathcal{I}^1(K))$ such that m is supported by every Baire set containing the ergodic measures and for each $f \in C(K)$,*

$$\mu(f) = \int_{\mathcal{I}^1(K)} \nu(f)\mathrm{d}m(\nu).$$

If, additionally, the ergodic measures form a Baire subset of $\mathcal{I}^1(K)$, m is unique.

Proof. The preceding propositions establish exactly the conditions needed to apply the Choquet-Bishop-De Leeuw and Choquet-Meyer theorems, which give the conclusion of the theorem by the characterisation of $\mathrm{Ext}(\mathcal{I}^1(K))$ as the ergodic measures. In particular, since the dual of $C(K)^*$ in the weak* topology consists of the evaluation maps (see Theorem IV.20 in [10]), m representing μ means that for every $f \in C(K)$,

$$\mu(f) = \hat{f}(\mu) = \int_{\mathcal{I}^1(K)} \hat{f}(\nu)\mathrm{d}m(\nu) = \int_{\mathcal{I}^1(K)} \nu(f)\mathrm{d}m(\nu),$$

where $\hat{f}$ denotes evaluation at f. For the uniqueness part we appeal to Corollary 6.8. □

The formulation of the above becomes even more elegant when K is also metrizable:

Theorem 7.10. *Let K be compact and metrizable and let $(T_i)_{i\in I} : K \to K$ be a family of commuting continuous maps. Then $\mathcal{I}^1(K) \neq \emptyset$, and for each $\mu \in \mathcal{I}^1(K)$ there exists a unique $m \in \mathcal{M}^1(\mathcal{I}^1(K))$ such that m is supported by the ergodic measures and for each $f \in C(K)$,*

$$\mu(f) = \int_{\mathcal{E}} \nu(f)\mathrm{d}m(\nu),$$

where $\mathcal{E}$ denotes the set of ergodic measures.

Proof. Just as in the proof of Lemma 3.6, Theorem 6.6 on p. 140 in [2] implies that $C(K)$ is separable. Furthermore, a Banach space is separable if and only if the closed unit ball of its dual is metrizable in the weak* topology. For a proof of this theorem we refer the reader to Theorem 5.1 on p. 134 of the same reference. Thus the closed unit ball in $C(K)^*$ is metrizable, and therefore the subset $\mathcal{I}^1(K)$ is also metrizable.

Since the compact convex set in question is now also metrizable, Choquet's existence and uniqueness theorems (Theorems 3.9 and 6.9) apply, yielding the conclusion. □

7.1 Concrete examples

The following example is originally due to Choquet and outlined on p. 77 in [8]:

Example 7.11. Invariant measures give rise to an easy construction of a Choquet simplex whose extreme points are not closed. Let $K = I \times S^1$ where I is the unit interval and S^1 is the unit circle, realized as $\{z \in \mathbb{C} \mid |z| = 1\}$. For notational ease, we will temporarily adopt the convention that $\exp(\cdot) = e^{2\pi i \cdot}$. Furthermore let $\lambda_{[0,1]}$ be the Lebesgue measure on the unit interval and let $T : K \to K$ be defined by $(t, \exp(s)) \mapsto (t, \exp(s + t))$.

Since I and S^1 are both compact and metrizable, so is K. Furthermore, it is not very difficult to see that T is continuous and bijective, so by the usual compact-Hausdorff argument, T is a homeomorphism. The theorem above then states that there exist invariant measures, and to each $\mu \in \mathcal{I}^1(K)$ there exists $m \in \mathcal{M}^1(\mathcal{I}^1(K))$ such that m is supported by the ergodic measures and m represents μ. Since K is metrizable, m is unique. We argue that in this case $\mathrm{Ext}(\mathcal{I}^1(K))$ is not closed, which is quite unintuitive when comparing to finite-dimensional simplices. It also showcases that Choquet's theorem is an improvement over the reformulated Krein-Milman theorem.

We define probability measures μ_n on K by $\mu_n = \sum_{k=0}^{n-1} 1/n\delta_{(1/n,\exp(k/n))}$. Since T is a bijection of the set $\{(1/n, \exp(k/n)) \mid 0 \leq k \leq n-1\}$ onto itself and $T(\delta_x) = \delta_{T(x)}$, we see that μ_n is invariant for every $n \in \mathbb{N}$. To show that μ_n is ergodic, let $A \in \mathbb{A}_{\mu_n}$. Then, since $\mu_n(A \triangle T^{-1}(A)) = 0$ and μ_n is discrete, each $(1/n, \exp(k/n))$ must be in either none or both of A and $T^{-1}(A)$. If none of them are in A, $\mu_n(A) = 0$. If one is in A, say for k_0, we also have $(1/n, \exp(k_0/n)) \in T^{-1}(A)$. But then

$$T(1/n, \exp(k_0/n)) = (1/n, \exp((k_0 + 1)/n)) \in A \cap T^{-1}(A),$$

which in turn implies $(1/n, \exp((k_0 + 2)/n) \in A \cap T^{-1}(A)$ and so forth. Therefore they must all be in $A \cap T^{-1}(A)$, that is, $\mu_n(A) = 1$. We conclude μ_n is ergodic and hence extreme in $\mathcal{I}^1(K)$.

We now claim $\mu_n \to \lambda_0$ in the weak* topology where λ_0 is the image measure on $\{0\} \times S^1$ defined by the map $[0, 1] \ni t \mapsto (0, \exp(t))$. Note that λ_0 is Radon as it coincides with the Haar measure on $\{0\} \times S^1$. We refer the reader to Chapter 11.1 in [3] for an exposition on Haar measures. By the usual characterisation of weak* convergence, we let $f \in C(K)$ and see that

$$\begin{aligned} |\mu_n(f) - \lambda_0(f)| &= \left| \sum_{k=0}^{n-1} \frac{1}{n} f(1/n, \exp(k/n)) - \int_{[0,1]} f \mathrm{d}\lambda_0 \right| \\ &= \left| \sum_{k=0}^{n-1} \int_{[0,1/n]} f(1/n, \exp(k/n)) - f(0, \exp(k/n + t))\mathrm{d}t \right| \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{k=0}^{n-1} \int_{[0,1/n]} |f(1/n, \exp(k/n)) - f(0, \exp(k/n + t))| \mathrm{d}t \\
&\leq \sum_{k=0}^{n-1} \frac{1}{n} \sup_{t \in [0,1/n]} \{|f(1/n, \exp(k/n)) - f(0, \exp(k/n + t))|\}.
\end{aligned}
$$

Since K is compact metrizable, f is uniformly continuous on K. Thus for large enough $n \in \mathbb{N}$ and independent of k, the above supremums can be made arbitrarily small. We conclude $\mu_n \to \lambda_0$ in the weak* topology.

It is then easy to see that λ_0 is not extreme. Indeed, if we let A be some Borel subset with $\lambda_0(A) = 1/2$, we may define probability measures $\mu_1(B) = 2\lambda_0(B \cap A)$ and $\mu_2(B) = 2\lambda_0(B \cap A^c)$ for $B \in \mathbb{B}(\{0\} \times S^1)$. But T restricted to $\{0\} \times S^1$ is just the identity, so of course μ_1 and μ_2 are invariant. Since $\mu_1 \neq \lambda_0$ and $\lambda_0 = 1/2(\mu_1 + \mu_2)$, λ_0 is not extreme.

While the above example showcased the even for Choquet simplices the set of extreme points need not have nice topological properties, the idea of non-closed extreme points has been taken even further. We conclude the thesis by mentioning the *Poulsen simplex*, which is is an extremely extraordinary simplex whose extreme points are actually *dense*. To fully appreciate its properties, we recall that a face F of a compact convex set X is a generalized exmtreme point, that is, if $x \in F$ and $x = \alpha y + (1-\alpha)z$ for $\alpha \in (0,1)$ and $y, z \in X$, then $y, z \in F$.

Example 7.12. Let $K = \{0,1\}^{\mathbb{Z}}$ be the set of doubly-infinite binary sequences and let $T : K \to K$ be the shift operator, that is, if $x = (x_k)_{k=-\infty}^{\infty} \in K$, T maps x_k to x_{k+1}. Then the set of invariant probability measures, $S = \mathcal{I}^1(K)$, is a realization of the Poulsen simplex. The Poulsen simplex has the following remarkable properties:

- Up to affine homeomorphism, S is the unique metrizable simplex such that $\overline{\mathrm{Ext}(S)} = S$.
- Any metrizable simplex X is affinely homeomorphic to a face of S.
- If F_1 and F_2 are closed proper faces of S and $\varphi : F_1 \to F_2$ is an affine homeomorphism, then φ extends to an affine homeomorphism $\bar{\varphi}$ of S onto itself.

The first and third property are contained in Theorem 2.3 in [5] while the second property and its proof is Theorem 2.5 in the same reference. The Poulsen simplex was first constructed by Ebbe T. Poulsen using ℓ^2, and we refer the reader to the original 1961 paper [9]. A proof of the fact that $\mathcal{I}^1(K)$ is a realization of the Poulsen simplex can be found on pages 618-619 in [4]. Theorem 3.9 in the same article states the fascinating fact that *any* compact metrizable simplex can be realized as $\mathcal{I}^1(K)$ for a compact metric space K and homeomorphism $T : K \to K$, a theorem originally due to T. Downarowicz.

References

[1] Gustave Choquet, *Lectures on Analysis: Representation Theory*, vol. 2., W.A Benjamin, Inc, 1969.

[2] John B. Conway, *A Course in Functional Analysis*, 2. ed., Springer, 2007.

[3] Gerald B. Folland, *Real Analysis*, 2. ed., John Wiley & Sons, INC, 1999.

[4] V.P. Fonf, J. Lindenstrauss, and R.R. Phelps, *Infinite Dimensional Convexity*, Handbook of the Geometry of Banach Spaces **1** (2001), 599–670.

[5] J. Lindenstrauss, G. Olsen, and Y. Sternfeld, *The Poulsen simplex*, Annales de l'institut Fourier **28** (1978), no. 1, 91–114.

[6] Jaroslav Lukeš, Jan Malý, Ivan Netuka, and Jiří Spurný, *Integral Representation Theory*, Walter de Gruyter GmbH & Co. KG, 2010.

[7] James Munkres, *Topology*, 2. ed., Pearson, 2014.

[8] Robert R. Phelps, *Lectures on Choquet's Theorem*, 2. ed., Springer, 2001.

[9] Ebbe T. Poulsen, *A simplex with dense extreme points*, Annales de l'institut Fourier **11** (1961), 83–87.

[10] Michael Reed and Barry Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, united kingdom ed., Academic Press Inc, 1980.

[11] Barry Simon, *Convexity: An Analytic Viewpoint*, Cambridge University Press, 2011.