ON CHOQUET'S THEOREM

YASSINE EL MAAZOUZ

Abstract. The main goal of this paper is to present Choquet's representation theorem and other similar results and to give a few examples of some interesting applications. We are not going to give proofs for every result (altough we give references to these) but we will rather focus on understanding the statements and proving the central theorems.

1. Introduction and background

A very powerful idea in convex and polyhedral geometry is the idea of representation (Minkowski's theorem): a point in a compact convex body in a finite dimensional vector space over $\mathbb{R}$ can be represented as barycentric combination of finitely many *extreme* points. This sort of result is very useful in many fields from convex optimization (especially linear programming) to probability theory which is going to be our main concern in what follows. It is then useful to start by restating the above result in probabilistic terms and we start with our first definition.

Definition 1.1. Let X be a non-empty compact subset of a locally convex space E and μ a probability measure on X. A point x in E is said to be represented by μ if we have $f(x) = \mu(f) := \int_X f d\mu$ for any linear function f on E. In this case x is also called the barycentrer of μ or the resultant of μ.

With this definition, Minkowski theorem translates to: a point in a compact convex body in a finite dimensional vector space over $\mathbb{R}$ can be represented by a probability measure with finite support in the the set of extreme points. There is a sharper version of this theorem due to Caratheodory which give a sharp upper bound on the number of extreme points that represent x. The use of a probability measure in this simple case seems very artificial but as we shall see later it will become a lot more natural.

The locally convex condition on E in definition 1.1 ensures that there are enough functionals in E^* so that at most one point can be represented by a probability measure μ. Notice that finding a measure that represents a point x is trivial since the Dirac measure δ_x does the job but the interesting part is can we find such a measure that is supported on the extreme points. Formally the problem at hand is the following:

Problem 1.2. Given a compact convex subset X of a locally convex space E, and a point $x \in X$, does there exist a probability measure μ on X supported by the extreme points of X which represents x? In case this measure exists is it unique?

Choquet [CM63] proved that the existence question has a positive answer in the case where X is metrizable while the uniqueness depends on the geometry of X. Allowing more general measures than Borel measures, Bishop-de Leeuw [BL59] have shown the existence of a representing measure without additional restrictions on X.

But before we get to see the theorems of Choquet and Bishop-de Leeuw we shall first translate the Riesz representation and the Krein-Milman theorems into our formalism and we will see later how Choquet and Bishop-de Leeuw theorems generalize the Krein-Milman theorem.

Let Y be a compact Hausdorff space and $C(Y)$ the Banach space of real valued continuous functions on Y (endowed with the supremum norm). Now let X be the set of continuous linear functions L on $C(Y)$ such that $L(\mathbf{1}) = 1$ where $\mathbf{1}$ is the constant function with value 1 on Y. The set X is compact convex in the locally convex space $E = C(Y)^*$ (the dual space of $C(Y)$, with weak* topology).

Theorem 1.3 (Riesz representation)**.** *For any $L \in X$ there exists a unique probability measure on Y such that $L(f) = \int_Y f d\mu$.*

To see where the extreme points are involved in the above statement, the injective map $y \mapsto (f \mapsto f(y))$ is a homeomorphism between the space Y and the set of extreme points of X so μ can be considered as a probability measure on X that gives mass 0 to the open set $X \setminus Y$ (we now think of Y as a subset of X via the injective homeomorphism above) so the support of μ is in the closed subset $Y \subset X$ which is in the set of extreme point of X.

In the above situation the extreme points of X form a compact Borel subset and the representation turned out to be unique. This does not happen in the general case and the first point is actually source of many problems in general (one can come up with examples where the set of extreme points is not closed even in finite dimension).

To reformulate the Krein-Milman theorem we first give a preliminary result in the form of

Proposition 1.4. *Suppose that Y is a compact subset of a locally convex space E. A point x in E is in the closed convex hull X of Y if and only if there exists a probability measure μ on Y which represents x.*

The Krein-Milman theorem states that: *If X is a compact convex subset of a locally convex space, then X is the closed convex hull of its extreme points.* Using Proposition 1.4 it is not very hard to reformulate the Krein-Milman theorem to the following equivalent form

Theorem 1.5 (Krein-Milman)**.** *Any point x of a compact convex subset X of a locally convex space can be represented by a probability measure supported on the closure of the set of extreme points of X.*

It becomes clear now that using a probability measure supported on only the extreme points (not their closure) would be a sharpening of the Krein-Milman theorem. Actually such a sharpening is necessary in many situations since as explained in [Kle59], "most" compact convex sets of infinite dimensional Banach spaces are the closure of their extreme points. So in these cases the Krein-Milman theorem is not very useful since one can always use Dirac masses to represent points.

As we have said above the main issue in finding a probability measure supported by extreme points is the nature of the set of extreme points itself which can fail to be a Borel set. In the case where X is metrizable this handicap can be overcome since X in this case would be G_δ set (countable intersection of open sets) according to

Proposition 1.6. *If X is a metrizable compact convex subset of a topological real vector space E then the set* $\operatorname{extr}(X)$ *of extreme points of X is a G_δ set.*

We recall that one can always use Dirac point masses to represent points and notice that if x is not an extreme point then it has more that one representation since it is the barycenter of at least two distinct points. As one might expect, the extreme points are exactly the points of X which admit only one representation (which is then the Dirac measure in the point itself) [Bau61].

To conclude the introduction we note an instance where the the Krein-Milman theorem yields interesting results.

Theorem 1.7 (Bernstein)**.** *If f is a bounded and completely monotone function on $(0, +\infty)$ then there exists a unique non-negative Borel measure μ on $[0, \infty]$ such that $\mu([0, \infty]) = f(0^+)$ and for each $x > 0$ we have*

$$f(x) = \int_0^\infty e^{-\alpha x} \mu(d\alpha)$$

Notice that the converse is true since any function that can be represented as above is necessarily completely monotone and also $\mu([0, \infty]) = f(0^+)$ follows by dominated convergence. The limit $f(0^+)$ exits by monotonicity (but might be ∞)

Proof sketch. We only give a sketch of a proof using the machinery we have introduced and we refer to [Phe01] for a complete proof. We denote by $\mathcal{C}$ the cone of completely monotonic functions with $f(0^+) < \infty$ and $\mathcal{K}$ the compact convex set of such functions with $f(0^+) \leq 1$. We reduce to the case $f \in \mathcal{K}$ by considering $f/f(0^+)$ when $0 \neq f \in \mathcal{C}$. The set $\mathcal{K}$ is convex in the space $E := C^\infty((0, \infty), \mathbb{R})$ which is locally convex in the topology of uniform convergence (of the function and all its derivatives) on compact subsets of $(0, +\infty)$. It remains to show that $\mathcal{K}$ is compact in this topology and the extreme points of $\mathcal{K}$ are exactly the functions $x \mapsto e^{-\alpha x}$, $0 \leq \alpha \leq +\infty$. Applying the Krein-Milman theorem will then allow us then to conclude. □

2. Choquet's representation theorem in the metrizabe case

In this section we prove Choquet's representation theorem for a metrizable X which is actually a special case of the more general Choquet-Bishop-de Leeuw theorem since the proof is not very long and will give us an opportunity to introduce some useful tools for the general case. In this section X is a compact convex subset of a locally convex space E.

Let h be a real valued function defined on a convex set C. The function is called *upper semi-continuous* if for each real λ, $\{x : f(x) < \lambda\}$ is open, while it is *lower semi-continuous* if $-h$ is upper semi-continuous. Let A be the set of all affine functions on X which is a subset of the Banach space $C(X)$ (endowed with the supremum norm) and contains the constant functions. Furthermore, A contains all functions of the forms $x \mapsto f(x) + r$ with $f \in E^*$ (recall that E^* is the dual of E), $r \in \mathbb{R}$ and $x \in X$ so that A contains enough functions to separate the points of X.

Definition 2.1. If f is a bounded function on X and $x \in X$, let $\bar{f}(x) := \inf\{h(x), h \in A \text{ and } h \geq f\}$.

The function $\bar{f}$ is calle the upper envelope of f, has the following nice properties which follows fairly easily from the definition.

(1) $\bar{f}$ is concave, bounded and upper semi-continuous (hence Borel measurable)

(2) $f \leq \bar{f}$ and if f is concave semi-continuous then $f = \bar{f}$.

(3) If f, g are bounded then $\overline{f+g} \leq \bar{f} + \bar{g}$ and $|\bar{f} - \bar{g}| \leq \|f - g\|$.

(4) $\overline{f+g} = \bar{f} + g$ if $g \in A$ and $\overline{rf} = r\bar{f}$ if $r > 0$.

Theorem 2.2. *(Choquet) Suppose that X is a metrizable convex subset of a locally convex space E and x_0 is an element of X. Then there exists a probability measure μ on X which represents x_0 and is supported by the extreme points of X.*

Proof. (Bonsall) Suppose that X is matrizable, $C(X)$ (and hence A) is separable. There exists then a sequence of functions h_n in A such that $\|h_n\| = 1$ and $(h_n)_{n\geq 0}$ is dense in the unit sphere of A; in particular it separates the points of X. Let $f = \sum 2^{-n} h_n^2$ this function is well defined since $C(X)$ is Banach and the series converges absolutely hence $f \in C(X)$ and f is strictly convex (if $x \neq y$ then $h_n(x) \neq h_n(y)$ for some n i.e the affine function h_n is not constant on $[x, y]$. Then h_n^2 is strictly convex on $[x, y]$ hence f is.) Let B denote the subspace $A + \mathbb{R}f$ of $C(X)$ generated by A and f. Now from properties (3) and (4) the functional p defined on $C(X)$ by $p(g) = \bar{g}(x_0) \quad (g \in C(X))$ is sub-additive and $p(rg) = rp(g)$. We define a linear functional on B by $h + rf \mapsto h(x_0) + r\bar{f}(x_0)$. We want to show that $h(x_0) + r\bar{f}(x_0) \leq (\overline{h + rf})(x_0)$ i.e the functional we defined on B is dominated by p. If $r \geq 0$ then $\overline{h + rf} = h + r\bar{f}$ while if $r < 0$ then $h + rf$ is concave then $\overline{h + rf} = h + rf \geq h + r\bar{f}$. By the Hahn-Banach theorem there exists a linear functional m on $C(X)$ such that $m(g) \leq \bar{g}(x_0)$ for all $g \in C(X)$ and $m(h + rf) = h(x_0) + r\bar{f}(x_0)$ for $h \in A$ and $r \in \mathbb{R}$. If $g \in C(X)$ and $g \leq 0$ then $m(g) \leq \bar{g}(x_0) \leq 0$, i.e m is non-positive on non-positive functions and hence is continuous. By the Riesz representation theorem, there exists a non-negative regular Borel measure μ on X such that $1 = m(\mathbf{1}) = \mu(\mathbf{1}) := \int_X d\mu$ so μ is a probability measure and $\mu(f) = m(f) = \bar{f}(x_0)$. Since $f \leq \bar{f}$ we have $\mu(f) \leq \mu(\bar{f})$ and for $h \in A$ and $h \geq f$ we have $h(x_0) = m(h) = \mu(h) \geq \mu(\bar{f})$ so we get $\mu(f) = \bar{f}(x_0) \geq \mu(\bar{f})$. This means that μ vanishes on the complement of the set $\mathcal{E} = \{x : f(x) = \bar{f}(x)\}$. It remains to show that $\mathcal{E}$ is contained in the set of extreme points of X. Indeed of $x = \frac{1}{2}(y + z)$ where y, z are distinct points in X then since f is strictly convex we have $f(x) < \frac{1}{2}(f(y) + f(z) \leq \frac{1}{2}(\bar{f}(y) + \bar{f}(z)) \leq \bar{f}(x)$. □

Actually we have $\mathcal{E} := \{x : f(x) = \bar{f}(x)\}$ exactly the set of extreme points of X.

Definition 2.3. If μ and λ are probability measures such that $\mu(f) = \lambda(f)$ for all $f \in A$ we write $\mu \sim \lambda$.

Proposition 2.4. *If f is a continuous function on the compact convex set X then for each $x \in X$ we have, $\bar{f}(x) = \sup\{\int f d\mu : \mu \sim \delta_x\}$. This implies that $\bar{f}(x) = f(x)$ if x is an extreme point in X.*

Proof sketch. The last implication follows from the characterization of extreme points in terms of representing measures. To show that first assertion we need to show that $f'(x) := \sup\{\int f d\mu : \mu \sim \delta_x\}$ coincides with $\bar{f}$. We can see from the definition that f' is concave. Next we show that f' is upper-semi-continuous and conclude using a similar argument than the one used to show property (2) that $f' \geq \bar{f}$. We have on the other hand for $h \in A$ with $h \geq f$ and for any $\mu \sim \delta_x$, $h(x) = \mu(h) \geq \mu(f)$ then $f' \leq \bar{f}$ and we conclude. Refer to [Phe01] for full details. □

Now that we have seen the special case where X is metrizable we are ready to dive into the general case which is going to occupy us in the next section.

3. The Choquet-Bishop-de Leeuw theorem

Suppose now that X is a non-metrizable compact convex subset of a locally convex space E. The extreme points of X do not in general form a Borel set [**?**] which makes it hard to makes sense of the phrase "probability measure supported on extreme points of X". We can in this case either drop the requirement that μ is a Borel measure and work on a different σ-ring or apart the definition of "supported by" for Borel measures. We take the second route and require that μ vanish on *Baire* subsets of X which contain no extreme points (Baire sets are the sets of the σ-ring generated by the compact G_δ sets).

Theorem 3.1. *(Choquet-Bishop-de Leeuw) Suppose that X is a compact convex subset of a locally convex space E and $x_0 \in X$. Then there exists a probability measure μ on X which represents x_0 and vanishes on every Baire subset of X which is disjoint from the set of extreme points of X.*

The proof of this theorem will occupy us for the rest of this section. We denote by A and C the set of affine and convex functions on X respectively. By $C - C$ we mean the space of function of the form $f - g$ with $f, g \in C$. This space is a lattice under the usual partial ordering since it is max-stable [by $\max(f_1 - g_1, f_2 - g_2) = \max(f_1 + g_2, f_2 + g_1) - (g_1 + g_2) \in C - C$]. Since $A \subset C - C$ the latter separates the points of X and contains constant functions and thus applying the Stone-Weierstrass theorem the space $C - C$ is dense in the norm topology of $C(X)$.

Definition 3.2. If λ and μ are non-negative regular Borel measures on X, write $\lambda \succ \mu$ if $\lambda(f) \geq \mu(f)$ for each f in C.

The relation $\succ$ defined above is clearly reflexive and transitive. The fact that it is anti-symmetric (i.e if $\lambda \succ \mu$ and $\mu \succ \lambda$ then $\lambda = \mu$) is a simple result of the density of $C - C$ in $C(X)$. Notice that for $f \in A$ then f and $-f$ are both in C so that $\lambda \succ \mu$ means $\mu(f) = \lambda(f)$. Also if $\mu \sim \delta_x$ then $\mu \succ \delta_x$ since $f \in -C$ (i.e f concave) then $f = \bar{f}$ and hence $f(x) = \inf\{h(x) : h \in A, h \geq f\} = \inf\{\mu(h) : h \in A, h \geq f\} \geq \mu(f)$ meaning that $\delta_x(-f) \leq \mu(-f)$ for all $-f \in C$.

We will be dealing with measures that are maximal for this partial ordering and such measures will be called maximal. The idea is that when $\lambda \succ \mu$ the support of λ is closer in some sense to the extreme points of X than that of μ. This can be checked on simple situations (take X to be a triangle in a 2-dimensional plane for example). We then hope that maximal measures with be supported on the extreme points in the sense that we discussed above.

Lemma 3.3. *If λ is a non-negative measure on X, then there exists a maximal measure μ such that $\mu \succ \lambda$.*

Proof. The proof uses Zorn's lemma and is omitted in this write-up. See [Phe01] for a detailed proof. □

Proposition 3.4. *If μ is a maximal measure on X, then $\mu(f) = \mu(\bar{f})$ for each continuous function f on X.*

Proof. Let $f \in C(X)$ and define the linear functional L on the one dimension subspace $\mathbb{R}f$ by $L(rf) = r\mu(\bar{f})$ and define sub-linear functional p on $C(X)$ by $p(g) = \mu(\bar{g})$. If

$r \geq 0$, then $L(rf) = p(rf)$ and of $r < 0$ then $0 = \overline{rf - rf} \leq \overline{rf} + \overline{-rf} = \overline{rf} - r\bar{f}$ hence $L(rf) = \mu(r\bar{f}) \leq \mu(\overline{rf}) \leq p(rf)$ then p dominates L on $\mathbb{R}f$. Then the Hahn-Banach theorem gives the existence of and extension L' of L to $C(X)$ such that $p \geq L'$. For $g \leq 0$ we have $\overline{g} \leq 0$ then $L'(g) \leq p(g) = \mu(\bar{g}) \leq 0$. Then $L' \geq 0$ and there exists then a non-negative measure ν on X such that $L'(g) = \nu(g)$ for each $g \in C(X)$. If g is convex then $-g$ is concave meaning that $\bar{-}g = -g$ so that $\nu(-g) \leq p(-g) = \mu(-g)$. Then means that if g is convex then $\mu(g) \leq \nu(g)$ hence $\nu \succ \mu$. Since μ is maximal we deduce that $\nu = \mu$ therefore $\mu(f) = \nu(f) = L(f) = \mu(\bar{f})$ which finishes the proof. □

The converse of the above result is actually also true [Phe01]. An important consequence of the above proposition is that: *If μ is a maximal measure, then μ is supported by the set $\{x : f(x) = \bar{f}(x)\}$ for each f in $C(X)$ (in particular in C)* and we have seen that each of these sets contains the set of extreme points of X. Now if C contains a strictly convex function f_0 we would have as in the proof of Choquet's theorem that $\text{extr}(X) = \{x : f_0(x) = \bar{f}_0(x)\}$ and the proof would be complete. However the existence of a strictly convex function actually implies that the set X is metrizable [Her61]. So the best we can hope for is that $\text{extr}(X)$ is the intersection of the sets $\{x : f(x) = \bar{f}(x)\}$ when $f \in C$. And indeed if $\bar{f}(x) = f(x)$ for each $f \in C$ and if $x = \frac{1}{2}(y + z)$ with $y, z \in X$ then

$$f(y) + f(z) \geq 2f(x) = 2\bar{f}(x) \geq \bar{f}(y) + \bar{f}(z) \geq f(y) + f(z)$$

meaning that $2f(x) = f(y) + f(z)$ for all $f \in C$ and thus for all $f \in C - C$ which is dense in $C(X)$ and hence $x = y = z$ i.e $x \in \text{extr}(X)$. It remains to tackle the harder task which is to show that maximal measures vanish on the Baire sets which are disjoint from $\text{extr}(X)$. For that it is enough to show that $\mu(D) = 0$ for any G_δ set disjoint from $\text{extr}(X)$ since $\mu(B) = \sup\{\mu(D) : D \subset B, D \text{ a compact } G_\delta\}$. We may assume that D is a compact subset of a G_δ set which is disjoint from $\text{extr}(X)$. To show that if $\mu(D) = 0$ we first use Urysohn's lemma to choose a non-decreasing sequence $\{f_n\}$ of continuous functions on X with $-1 \leq f_n \leq 0$, $f_n \equiv -1$ on D and $\lim_n f_n(x) = 0$ for $x \in \text{extr}(X)$. We show that if μ is maximal then $\lim_n \mu(f_n) = 0$ this will imply that $\mu(D) = 0$. We shall need a couple of preliminary results to get the desired "limit" result.

The first of the following couple of lemma reduces the problem to the Choquet theorem on metrizable X. We will make use of the fact that for each $x \in X$ there exists $\mu \sim \delta_x$ supported on $\text{extr}(X)$. Since we can't in general extend a element $f \in A$ to an element in E^* this is stronger than stated version of Choquet's theorem above. We shall not give a proof of these preliminary results and we refer to [Phe01] for detailed proofs and discussion.

Lemma 3.5. *Suppose that $(f_n)_n$ is a bounded sequence of concave upper semi-continuous functions on X, with $\liminf f_n(x) \geq 0$ for each $x \in \text{extr}(X)$. Then $\liminf f_n \geq 0$ on X*

Proof. We refer the reader to [Phe01] for a detailed proof. □

Now this first result is not very unexpected and one can see how it works in finite dimension although its much less trivial in the general case.

Lemma 3.6. *If μ is a maximal measure of X and if (f_n) is a non-decreasing sequence in $C(X)$ such that $1 \leq f_n \leq 0$ for all n with $\lim f_n(x) = 0$ for $x \in \operatorname{extr}(X)$, then* $\lim \mu(f_n) = 0$

Proof. We refer the reader to [Phe01] for a detailed proof. □

These two results alongside the previous discussion actually show that: *any maximal measure μ vanishes on any G_δ set of X that is disjoint of* $\operatorname{extr}(X)$. As a result, any maximal measure is supported by any closed set containing $\operatorname{extr}(X)$ and thus the Choquet-Bishop-de Leeuw theorem generalizes the Krein-Milman theorem stated above.

Finally we give a different statement of the Choquet-Bishop-de Leeuw theorem which is more practical for applications.

Theorem 3.7. *(Bishop-de Leeuw) Suppose that X is a compact convex subset of a locally convex space and let $\mathcal{S}$ the ring σ-ring of subsets of X which is generated by* $\operatorname{extr}(X)$ *and the Baire sets. Then for each point x_0 in X there exists a non-negative measure μ on $\mathcal{S}$ with $\mu(X) = 1$ such that μ represents x_0 and $\mu(\operatorname{extr}(X)) = 1$.*

The proof [Phe01] is not very difficult once we have the Choquet-Bishop-de Leeuw theorem in our hands.

As we've seen so far the idea of representation by means of extreme points, although it is very intuitive and unsurprising in Euclidean spaces, is quite tricky in its full generality. The arguments we went through (or referred to in other sources) required a combination of some highly non-trivial results in functional analysis and measure theory. One has to appreciate the power of this circle of ideas since it simplifies a great deal of arguments (for instance to prove a statement one might end up only having to prove it for extreme points since all other points are a "convex combination" of these). The nice thing in particular about Choquet-Bishop-de Leeuw is the generality of the statement which only requires compactness and convexity of X and E to be locally convex and once could be surprised how diverse the situation is which this can be applied.

References

[Bau61] Heinz Bauer. Silovscher rand und dirichletsches problem. In *Annales de l'institut Fourier*, volume 11, pages 89–136, 1961.

[BL59] Errett Bishop and Karel De Leeuw. The representations of linear functionals by measures on sets of extreme points. In *Annales de l'institut Fourier*, volume 9, pages 305–331, 1959.

[CM63] Gustave Choquet and Paul-André Meyer. Existence et unicité des représentations intégrales dans les convexes compacts quelconques. In *Annales de l'institut Fourier*, volume 13, pages 139–154, 1963.

[Her61] Michel Hervé. Sur les representations integrales a l'aide des points extremaux dans un ensamble compact convexe metrisable. *CR Acad. Sci. Paris*, 253:366–368, 1961.

[Kle59] Victor Klee. Some new results on smoothness and rotundity in normed linear spaces. *Mathematische Annalen*, 139(1):51–63, 1959.

[Phe01] Robert R Phelps. *Lectures on Choquet's theorem*. Springer Science & Business Media, 2001.

Email address: `yassine.el-maazouz@berkeley.edu`