**Riesz representation theorem**

Notes by Alex Furman.

Let $X$ be a compact metrizable space. Denote by $C(X)$ the vector space of all continuous functions $f : X \to \mathbb{R}$, and for $f \in C(X)$ denote

$$
\|f\|_u = \max_{x \in X} |f(x)|.
$$

The max operation is well defined, because any continuous functions on a compact space is bounded and attains its maximum. Furthermore, $\|f\|_u$ is a norm on $C(X)$, namely

- $\|f\|_u \geq 0$ and $\|f\|_u = 0$ iff $f \equiv 0$,
- $\|c \cdot f\|_u = |c| \cdot \|f\|_u$ for any scalar $c$,
- $\|f + g\|_u \leq \|f\|_u + \|g\|_u$.

Finally observe that $C(X)$ is complete with respect to this norm: any Cauchy sequence has a limit in $C(X)$. This is a reformulation of the statement that a uniform limit of continuous functions is continuous.

Recall that a finite, regular, Borel measure on $X$, is a finite (positive) measure defined on the Borel $\sigma$-algebra $\mathcal{B}_X$ of $X$, and satisfying for every $E \in \mathcal{B}_X$

$$
\mu(E) = \sup\{\mu(K) \mid K \subseteq E, K \text{ is compact}\} = \inf\{\mu(U) \mid E \subseteq U, U \text{ is open}\}.
$$

Denote by $M_+(X)$ the collection of all finite, positive, regular, Borel measures. A signed, regular, Borel measure, is a difference between two finite, positive regular, Borel measures; we denote by $M(X)$ the collection of all such signed measures.

By Jordan’s decomposition any $\mu \in M(X)$ has a unique decomposition

$$
\mu = \mu_+ - \mu_-, \quad \mu_+ \perp \mu_-, \quad \mu_+ + \mu_- \in M_+(X).
$$

Recall, that any $\mu \in M(X)$ can be written as $\mu = \mu_1 - \mu_2$ with $\mu_1, \mu_2 \in M_+(X)$ in many ways. However, in all such presentations one has $\mu_1 \geq \mu_+$ and $\mu_2 \geq \mu_-$ (in the sense that $\mu_1(E) \geq \mu_+(E)$ for all $E \in \mathcal{B}_X$).

The total variation $\|\mu\|$ of $\mu \in M(X)$ is given by

$$
\|\mu\| = \|\mu\|_X = \mu_+(X) + \mu_-(X).
$$

**Claim 1.** The space $M(X)$ is a vector space, and $\|\mu\|$ is a norm on $M(X)$.

Let us show the triangle inequality. Let $\mu, \nu \in M(X)$ and denote $\eta = \mu + \nu$. Let $\eta = \eta_+ - \eta_-$ be the Jordan decomposition, and denote $\eta_1 = \mu_+ + \nu_+$, and $\eta_2 = \mu_- + \nu_-$. Then $\eta_1 \leq \eta_1$ and $\eta_2 \leq \eta_2$. Therefore

$$
\|\mu + \nu\| = \eta_+(X) + \eta_-(X) \leq \eta_1(X) + \eta_2(X)
$$

$$
= \mu_+(X) + \nu_+(X) + \mu_-(X) + \nu_-(X) = \|\mu\| + \|\nu\|.
$$

The aim of these notes is to sketch the proof of the following:

**Theorem 2** (Riesz). The dual $C(X)^*$ to $C(X)$ is isometrically isomorphic to $M(X)$ equipped with the total variation norm, where $\mu \in M(X)$ corresponds to the functional $\Lambda_\mu$ on $C(X)$ defined by

$$
\Lambda_\mu(f) = \int_X f \, d\mu = \int_X f \, \mu_+ - \int_X f \, \mu_-.
$$
Since continuous functions are Borel measurable, the last integrals are defined; they evaluate to finite quantities because \( f \) is bounded and \( \mu_{\pm} \) are finite measures. Since integration is linear, one has

\[
\Lambda_{\mu}(f + g) = \Lambda_{\mu}(f) + \Lambda_{\mu}(g), \quad \Lambda_{\mu}(c \cdot f) = c \cdot \Lambda_{\mu}(f).
\]

Furthermore, since for every \( f \in C(X) \), one has \( |f(x)| \leq \|f\|_u \), we have

\[
|\Lambda_{\mu}(f)| = \left| \int f \, d\mu_+ - \int f \, d\mu_- \right| \leq \int |f| \, d\mu_+ + \int |f| \, d\mu_- \leq \|f\|_u \cdot \mu_+(X) + \|f\|_u \cdot \mu_-(X)
\]

Hence \( \Lambda_{\mu} \) is continuous, and has norm

\[
\|\Lambda_{\mu}\| = \sup_{\|f\|_u = 1} |\Lambda_{\mu}(f)| \leq \|\mu\|.
\]

For any \( f \in C(X) \) and \( \mu, \nu \in M(X) \) and scalar \( c \)

\[
\Lambda_{\mu + \nu}(f) = \Lambda_{\mu}(f) + \Lambda_{\nu}(f), \quad \Lambda_{c \cdot \mu}(f) = c \cdot \Lambda_{\mu}(f).
\]

Thus the map \( \mu \mapsto \Lambda_{\mu} \) is a linear map \( M(X) \to C(X)^* \). We first note that it is isometric (with respect to the dual norm on \( C(X)^* \) and the total variation norm on \( M(X) \)). In particular, the map \( M(X) \to C(X)^* \) is injective.

**Claim 3.** \( \|\Lambda_{\mu}\| = \|\mu\| \).

**Proof.** Since \( \|\Lambda_{\mu}\| \leq \|\mu\| \) is already known, we need to obtain the reverse. Given \( \epsilon > 0 \) we will find \( f \in C(X) \) with \( \|f\|_u = 1 \) with

\[
\Lambda_{\mu}(f) > \|\mu\| - \epsilon.
\]

Since \( \mu = \mu_+ - \mu_- \) with \( \mu_+ \perp \mu_- \), there is a partition \( X = X_1 \sqcup X_2 \) into Borel sets with

\[
\mu_+(X \setminus X_1) = 0, \quad \mu_-(X \setminus X_2) = 0.
\]

By regularity, there are compact sets \( K_1 \subset X_1 \) and \( K_2 \subset X_2 \) so that

\[
\mu_+(X \setminus K_1) < \frac{\epsilon}{4}, \quad \mu_-(X \setminus K_2) < \frac{\epsilon}{4}.
\]

Next observe that there exists a continuous function \( f \) on \( X \) with

\[
f|_{\mu_1} = 1, \quad f|_{\mu_2} = -1, \quad -1 \leq f(x) \leq 1 \quad (x \in X).
\]

Indeed, fix some metric \( d \) on \( X \) and define

\[
f(x) = \frac{d(x, K_2) - d(x, K_1)}{d(x, K_1) + d(x, K_2)} \quad \text{where} \quad d(x, A) = \inf_{y \in A} d(x, y).
\]

For such a function \( f \) one has \( \|f\|_u = 1 \) and

\[
\Lambda_{\mu}(f) = \int_{X_1} f \, d\mu_+ - \int_{X_2} f \, d\mu_-
\]

\[
\geq 1 \cdot \mu_+(K_1) - (1) \cdot \mu_-(K_2) - \mu_+(X_1 \setminus K_1) - \mu_-(X_2 \setminus K_2)
\]

\[
\geq \mu_+(X_1) - \frac{\epsilon}{4} - \mu_-(X_2) - \frac{\epsilon}{4} + 2 \cdot \frac{\epsilon}{4} = \|\mu\| - \epsilon.
\]

\( \square \)
The main point of Riesz’ theorem is that the isometric imbedding
\[ M(X) \rightarrow C(X)^*, \quad \mu \mapsto \Lambda_\mu \]
is surjective. Namely that any continuous linear functional \( \Lambda \) on \( C(X) \) is \( \Lambda = \Lambda_\mu \) for some (necessarily unique) signed measure \( \mu \in M(X) \).

The next claim reduces the problem to showing this for \textit{positive functionals}. We say that a functional \( \Lambda \in C(X)^* \) is \textit{positive} if \( \Lambda(g) \geq 0 \) for non-negative continuous \( g \) on \( X \).

**Claim 4.** For any \( \Lambda \in C(X)^* \) there exist positive functionals \( \Lambda', \Lambda'' \in C(X)^* \), so that
\[ \Lambda(f) = \Lambda'(f) - \Lambda''(f) \quad (f \in C(X)). \]

\textit{Sketch of the proof.} First consider the cone of positive functions \( f \geq 0 \) (short for continuous functions \( f : X \rightarrow [0, \infty) \)). For \( f \geq 0 \) define
\[ \Lambda'(f) = \sup\{\Lambda(g) \mid 0 \leq g \leq f\}. \]

For all functions \( f, f_1, f_2 \geq 0 \) and scalars \( c \geq 0 \) show
\[ \Lambda'(c \cdot f) = c \cdot \Lambda'(f), \quad \Lambda'(f_1 + f_2) = \Lambda'(f_1) + \Lambda'(f_2). \]
The first fact is obvious. For the second, the inequality
\[ \Lambda'(f_1 + f_2) \geq \Lambda'(f_1) + \Lambda'(f_2) \]
is direct. To show ” \leq ” consider arbitrary \( 0 \leq g \leq f_1 + f_2 \) and set
\[ g_1 = \min(g, f_1), \quad g_2 = g - g_1. \]
Then \( 0 \leq g_1 \leq f_1 \) and \( 0 \leq g_2 \leq f_2 \), and \( \Lambda'(f_1 + f_2) \leq \Lambda'(f_1) + \Lambda'(f_2) \) follows. One then extends \( \Lambda' \) to all of \( C(X) \) by setting
\[ \Lambda'(f) = \Lambda'(f^+) - \Lambda'(f^-) \]
and checks, arguing as in the development of the Lebesgue integral from the integral for positive functions, that \( \Lambda' \) is a positive linear functional. It then follows that the functional \( \Lambda'' = \Lambda - \Lambda' \) is positive too. See [1, Lemma 7.15] for details. \( \square \)

Hence, to prove Riesz theorem it remains to show that

**Theorem 5** (Special case of Riesz representation theorem). \textit{Let \( \Lambda \) be a positive linear functional on \( C(X) \), then there exists a positive \( \mu \in M_+(X) \) so that}
\[ \Lambda(f) = \int_X f \, d\mu. \]

**Claim 6.** **Theorem 5** holds for the case where \( X \) is a Cantor set.

\textit{Proof.} We start with some remarks about Cantor sets.

Consider the following construction. Let \( F_1, F_2, \ldots \) be a sequence of finite sets with \( |F_n| \geq 2 \) elements each. Consider the space
\[ K = \prod_{n=1}^\infty F_n \]
of sequences \( x = (x_n)_{n=1}^\infty \) where \( x_n \in F_n \) for each \( n \in \mathbf{N} \). Define a metric \( d \) on \( K \) by setting
\[ d(x, y) = 2^{-n} \quad \text{if} \quad x_1 = y_1, \ldots, x_n = y_n, \quad x_{n+1} \neq y_{n+1}. \]
for \( x \neq y \), and \( d(x, x) = 0 \). Then \( d \) is a metric (check). In fact, it satisfies
\[
d(x, z) \leq \max(d(x, y), d(y, z)) \quad (x, y, z \in K).
\]
Furthermore \( K \) is compact space with respect to \( d \). One can show, that up to homeomorphism, \( K \) does not depend on the choice of the basic finite sets \( \{F_n\}_{n=1}^{\infty} \).

The usual, middle third Cantor sets, is easily seen to be homeomorphic to the infinite product of two-point sets: \( F_n = \{0, 1\} \) for all \( n \in \mathbb{N} \), via
\[
h : \{0, 1\}^\mathbb{N} \to C \subset [0, 1], \quad h(x) = \sum_{n=1}^{\infty} \frac{2x_n}{3^n}.
\]
Since all such spaces are homeomorphic, one talks about the Cantor set, as an abstract topological space. Such spaces admit a nice characterization: any compact totally disconnected space without isolated points is homeomorphic to the Cantor set.

Consider sets of the form
\[
A = S \times F_{k+1} \times F_{k+2} \times \cdots \subset K
\]
where \( k \) is some integer and \( S \subset F_1 \times \cdots \times F_k \) is an arbitrary subset. Let \( \mathcal{A} \) be the collection of all such sets. Then \( \mathcal{A} \) is an algebra (check !), and each \( A \in \mathcal{A} \) is both open and closed (hence compact) subset of \( K \). Therefore their characteristic functions \( f = \chi_A \) are continuous (!) on \( K \). Define \( \rho : \mathcal{A} \to [0, \infty) \) by
\[
\rho(A) = \Lambda(\chi_A).
\]
We claim that \( \rho \) is a pre-measure. Indeed, suppose \( A = \bigsqcup_{i=1}^{\infty} A_i \), where \( A_i \in \mathcal{A} \) and \( A \in \mathcal{A} \). All sets in \( \mathcal{A} \) are both compact and open. Hence one may view \( A = \bigsqcup_{i=1}^{\infty} A_i \) as an open cover of a compact set. Hence it contains a finite subcover. But the cover consisted of disjoint sets. This means that all, but finitely many \( A_i \) are empty sets. So it remains to check finite additivity of \( \rho \):
\[
\rho(\bigsqcup_{i=1}^{N} A_i) = \Lambda(\biguplus_{i=1}^{N} \chi_{A_i}) = \Lambda(\biguplus_{i=1}^{N} \chi_{A_i}) = \sum_{i=1}^{N} \Lambda(\chi_{A_i}) = \sum_{i=1}^{N} \rho(A_i).
\]
Hence, by Caratheodory theorem there is a unique extension \( \mu \) of \( \rho \) to the \( \sigma \)-algebra, generated by \( \mathcal{A} \). Since cylinder sets is a base for the topology of \( K \), \( \mu \) is a Borel measure. It is regular because it corresponds to an outer measure defined by \( \rho \).

Finally, the formula
\[
\Lambda(f) = \int_{K} f \, d\mu
\]
is true for characteristic functions of cylinder sets and finite linear combinations of such functions. Since the latter span a dense subspace of \( C(X) \), it follows that \( \Lambda = \Lambda_{\mu} \).

**Claim 7.** For any compact metrizable space \( X \) there is a continuous surjective map \( p : K \to X \) from a Cantor set \( K \).

**Proof.** Let \( d \) be some metric on \( X \) (defining the given topology). We denote by \( B_{x,r} = \{y \in X \mid d(x, y) < r \} \) the closed ball of radius \( r \) around \( x \in X \). We shall define finite sets \( F_1, F_2, \ldots \) and maps
\[
p_k : F_1 \times \cdots \times F_k \to X
\]
inductively as follows. Since $X$ is compact, it can be covered by finite number $N_1$ of balls of radius 1. Let $F_1 = \{1, \ldots, N_1\}$ and $p_1 : F_1 \to X$ be a map so that

$$X = \bigcup_{x \in F_1} B_{p_1(x), 1}. $$

For each $x \in F_1$, the ball $B_{p_1(x), 1} \subset X$ has compact closure. Hence for some $N_2$ one can find a map $p_2 : F_1 \times F_2 \to X$ so that for $x_1 \in F_1, x_2 \in F_2$

$$p_2(x_1, x_2) \in B_{p_1(x_1), 1} \subset \bigcup_{y \in F_2} B_{p_2(x_1, y), y}. $$

One continues by induction to find finite sets $F_k$ and maps $p_k : F_1, \ldots, F_k \to X$ so that

$$p_{k+1}(x_1, \ldots, x_{k+1}) \in B_{p_k(x_1, \ldots, x_k), 1} \subset \bigcup_{y \in F_{k+1}} B_{p_{k+1}(x_1, \ldots, x_k, y), y}, $$

We define $K = \prod_{n=1}^\infty F_n$ and a map $p : K \to X$ by

$$p(x) = \lim_{n \to \infty} p_n(x_1, \ldots, x_n).$$

Observe that this is a limit of a Cauchy sequence, hence the limit exists. It is continuous, because $d(p(x), p(y)) \leq 1/n$ provided $x_1 = y_1, \ldots, x_{n+1} = y_{n+1}$, i.e. $d_K(x, y) < 2^{-n+1}$. Finally, $p(K) = X$ because the image is dense in $X$.

**Proof of Riesz representation theorem 2.** We shall deduce the result from the special case of the Cantor set, using Hahn-Banach theorem on extension of linear functionals (see Thm 5.6 in Folland). Let $\Lambda \in C(X)^*$ be a positive functional. Let $p : K \to X$ be a surjective continuous map from a Cantor set $K$ (Claim 7). It defines an isometric linear imbedding

$$C(X) \to C(K) \quad f \mapsto f \circ p.$$ 

Thus $\Lambda$ may be viewed as a linear functional defined on a closed subspace

$$C(X) \subset C(K).$$

We use Hahn-Banach theorem (cf. [1, Theorem 5.6]) to extend $\Lambda$ on $C(X)$ to a continuous linear functional $\hat{\Lambda}$ on $C(K)$. Using Claim 4 we may view $\hat{\Lambda}$ as a difference of positive functionals

$$\hat{\Lambda} = \hat{\Lambda}' - \hat{\Lambda}''$$

on $C(K)$. By Claim 6 these functionals are given by positive measures $\hat{\mu}', \hat{\mu}'' \in M_+(K)$; hence $\hat{\Lambda}$ corresponds to $\hat{\mu} = \hat{\mu}' - \hat{\mu}'' \in M(K)$.

Finally, let $\mu \in M(X)$ be the push-forward of $\hat{\mu} \in M(K)$ by $p : K \to X$, namely

$$\mu(E) = \hat{\mu}(p^{-1}(E)), \quad \int_X f \, d\mu = \int_K f \circ p \, d\hat{\mu} \quad (E \in \mathcal{B}_X, f \in C(X)).$$

Since $\hat{\mu}$ defines $\hat{\Lambda}$, which extends $\Lambda$ from $C(X)$, it follows that

$$\Lambda(f) = \hat{\Lambda}(f \circ p) = \int_K f \circ p \, d\hat{\mu} = \int_X f \, d\mu \quad (f \in C(X)).$$

\[\square\]
Further remarks

**Locally compact case** Riesz’ representation theorem applies to a more general situation of locally compact space $X$, where the Banach space $C(X)$ is replaced by $C_0(X) –$ the space of all continuous functions $f : X \to \mathbb{R}$ with $|f(x)| < \epsilon$ for $x$ outside sufficiently large compact subset $K_\epsilon \subset X$. Then $C_0(X)$ is a Banach space with respect to the uniform norm $\|f\|_u$, and $C_0(X)^*$ is still identified with $M(X)$. One can deduce this result from the compact case, by looking at a one point compactification $\bar{X} = X \cup \{\infty\}$ of $X$, which identifies $C_0(X)$ with the codimension one subspace of $C(\bar{X})$ consisting of functions vanishing at $\infty$, and $M(X)$ can be viewed as the restriction of $M(\bar{X})$ to $X$.

**Weak-* convergence** of measures. There is a topology on $M(X)$ which is weaker than the topology defined by the total variation. One says that $\mu_n \to \mu$ in weak-* topology if for all $f \in C(X)$:

$$\int f \, d\mu_n \to \int f \, d\mu.$$

**Theorem 8.** If $X$ is a metrizable compact, then every closed ball $\bar{B}_R = \{\mu \in M(X) \mid \|\mu\| \leq R\}$ is weak-* compact.

Sketch of the proof, as an advanced exercise:

1. Show that $C(X)$ is separable, i.e., contains a countable set $\{f_j\}_{j=1}^\infty$, which is dense with respect to $\|f\|_u$.
2. Given a sequence $\mu_n \in B$ extract a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ so that

   $$a_j = \lim_{k \to \infty} \int f_j \, d\mu_{n_k}$$

   exist for each $j = 1, 2 \ldots$ (use diagonal construction).
3. Define $\Lambda(f_j) = a_j$ and observe that $\Lambda$ extends to a continuous linear functional on $C(X)$ with norm $\|\Lambda\| \leq R$ (note that $|a_j| \leq R \cdot \|f_j\|$).
4. Applying Riesz’ theorem, deduce that $\Lambda$ corresponds to some $\mu \in \bar{B}_R \subset M(X)$, and show that $\mu_{n_k} \to \mu$ in weak-* topology.

The space of probability measures on $X$ is $P(X) = \{\mu \in M_+(X) \mid \mu(X) = 1\}$.

**Exercise 9.** The map $X \to P(X)$ given by $x \mapsto \delta_x$ is a homeomorphism onto its image in the weak-* topology. However, the image of this map $\{\delta_x \mid x \in X\}$ is discrete in the total variation norm; in fact $\|\delta_x - \delta_y\| = 2$ whenever $x \neq y \in X$.

**References**