Combinatorics in Banach space theory

Lecture 10

10 Steinitz's lemma

This section is devoted to a beautiful result in combinatorial geometry which is nowadays known as Steinitz's lemma. It was proved (in a slightly different version than the one presented below) by Steinitz in 1913 and was also considered by Lévy in 1905 who, however, gave an incomplete proof. Here, we address the approach by Bárány and Grinberg [BG06]. In subsequent sections we will discuss several striking applications of Steinitz's lemma, the first of whom will be a stability version of the classical Lyapunov theorem on the range of a vector measure, due to V. Kadets [Kad91]. In the sequel, we will also report on Beck's work [Bec62] and see how meaningful Steinitz's lemma is in the context of B-convexity and vector analogs of the strong law of large numbers.

Lemma 10.1 (Steinitz's lemma as presented by Bárány and Grinberg, 1981)**.** *Let $\|\cdot\|$ be any norm in the linear space $\mathbb{R}^d$, where $d \in \mathbb{N}$. Then, for every finite set of vectors $x_1, \ldots, x_n \in \mathbb{R}^d$ satisfying $\|x_j\| \leqslant 1$ for every $1 \leqslant j \leqslant n$ there exists a sequence of signs $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$ such that*

$$\Big\| \sum_{j=1}^{n} \varepsilon_j x_j \Big\| \leqslant d.$$

Moreover, in the case where $\|\cdot\|$ is the Euclidean norm the right-hand side may be decreased to $\sqrt{d}$.

Proof. The key idea is to consider the collection of linear dependencies for the vectors $x_1, \ldots, x_n$, that is, the set

$$\mathcal{V} = \Big\{ \alpha = (\alpha_1, \ldots, \alpha_n) \in [-1,1]^n \colon \sum_{j=1}^{n} \alpha_j x_j = 0 \Big\}.$$

Plainly, $\mathcal{V}$ forms a convex polytope contained in the unit cube of $\mathbb{R}^n$, whose dimension equals at least $n-d$. We claim that for every extreme point (vertex) α^* of $\mathcal{V}$ at least $n-d$ coordinates $\alpha_1^*, \ldots, \alpha_n^*$ are equal to ± 1. Indeed, for any $\alpha_{i_1}^*, \ldots, \alpha_{i_k}^* \in (-1,1)$ the vectors $x_{i_1}, \ldots, x_{i_k}$ are linearly independent because otherwise $\gamma_{i_1} x_{i_1} + \ldots + \gamma_{i_k} x_{i_k} = 0$, where all $|\gamma_{i_j}|$ may be arbitrarily small and not all of them are zero, so we may add such a combination to $\sum_{j=1}^{n} \alpha_j^* x_j$ keeping the coefficients inside $(-1,1)$, which is impossible as α^* is an extreme point. Consequently, every extreme point of $\mathcal{V}$ has at most d coordinates lying inside $(-1,1)$, thus at least $n-d$ of them are equal to ± 1. Now, fix any extreme point $\alpha^* \in \mathcal{V}$ and assume, with no loss of generality, that $\alpha_1^*, \ldots, \alpha_k^*$ (with $0 \leqslant k \leqslant d$) are all its coordinates that belong to $(-1,1)$. Define $u = \sum_{j=k+1}^{n} \alpha_j^* x_j$. This is just a part of the desired combination of $x_1, \ldots, x_n$ with ± 1 coefficients. We shall find the remaining k coefficients.

Consider the parallelogram

$$\mathcal{Q} = \left\{ \sum_{j=1}^{k} \beta_j x_j \colon \beta_j \in [-1,1] \text{ for each } j \in [k] \right\};$$

it is k-dimensional, since $x_1, \dots, x_k$ are linearly independent. Moreover, all sides of $\mathcal{Q}$ have length at most 2 and, according to how u was defined, $0 \in u + \mathcal{Q}$. Hence, to finish the proof we shall show the following geometric claim: If a point a belongs to a parallelogram $\mathcal{Q}$ (like the one above) spanned by k linearly independent vectors $v_1, \dots, v_k$ with $\|v_j\| \leqslant 1$ for each $j \in [k]$, then there exists at least one vertex of $\mathcal{Q}$ at distance at most k from a. This will do the job, since $k \leqslant d$.

For $k = 1$ our claim is obvious, so suppose that $k \geqslant 2$ and the claim holds true for $k-1$. If a lies on the boundary of $\mathcal{Q}$, then the assertion follows directly from the inductive hypothesis. So, suppose a lives in the interior of $\mathcal{Q}$. Then, there exists the least positive number r such that the ball centred at a and with radius r meets at least one of the faces of $\mathcal{Q}$ at some point, say b. Of course, $r = \|a - b\| \leqslant 1$ because otherwise $\mathcal{Q}$ would contain a ball with diameter greater than 2 which is impossible. The face of $\mathcal{Q}$ that contains b is itself a $(k-1)$-dimensional parallelogram, whence by our inductive hypothesis there is a vertex w of that face (being also a vertex of $\mathcal{Q}$) such that $\|b - w\| \leqslant k - 1$. Consequently, we have $\|a - w\| \leqslant \|a - b\| + \|b - w\| \leqslant 1 + (k-1) = k$.

Finally, observe that for the Euclidean norm a and b must be perpendicular and then a similar inductive argument and Pythagorean's theorem give the estimate $\|a - w\| \leqslant \sqrt{k} \leqslant \sqrt{d}$. □

Remark 10.2. The constant d is optimal in the general case which is seen by taking $x_j = e_j$, for $j \in [d]$, in the space ℓ_1^d. Also $\sqrt{d}$ is the best possible constant for the Euclidean norm which follows by defining $x_1, \dots, x_d$ to be any orthonormal set in $\mathbb{R}^d$. For the space ℓ_∞^d the assertion of Steinitz's lemma holds true with $6\sqrt{d}$. This follows from a result obtained by Spencer [Spe85] and, independently, by Gluskin [Glu88], which says that every point lying in a parallelogram $\mathcal{Q} \subset \ell_\infty^d$ (like the one in the above proof) is at distance at most $6\sqrt{d}$ from some of its vertices.

For some other versions of Steinitz's lemma and its several applications, the reader may consult Chapter 2 in [KK97].

11 A quantitative version of Lyapunov's convexity theorem

In 1940, Lyapunov [Lya40] showed that the range of any σ-additive, non-atomic vector measure with values in a finite-dimensional normed space is convex and compact. For warming up, we present below the proof of the one-dimensional version of Lyapunov's theorem which, being much simpler than the multidimensional one, is still quite non-trivial. It is not only interesting for its own sake, but will be used in the proof of Lemma 11.7 below. Recall that an *atom* of a given non-negative measure μ is any measurable set A with $\mu(A) > 0$ and such that for any its measurable subset B we have $\mu(B) \in \{0, \mu(A)\}$. We shall start with a simple lemma.

Lemma 11.1. *Let Σ be a σ-algebra and $\mu \colon \Sigma \to [0, \infty)$ be a non-atomic, σ-additive measure. Then for every $A \in \Sigma$ and every $\varepsilon > 0$ there exists a set $B \subset A$, $B \in \Sigma$, such that $0 < \mu(B) < \varepsilon$.*

Proof. By the assumption, we have $A = A_1 \cup B_1$, where both of the sets A_1 and B_1 are Σ-measurable and of positive measure. Hence, one of them, say A_1, has measure at most $\frac{1}{2}\mu(A)$. Repeating this argument for A_1 instead of A we get a sequence $(A_n)_{n=1}^{\infty}$ of Σ-measurable subsets of A such that $0 < \mu(A_n) \leqslant 2^{-n}\mu(A)$ for each $n \in \mathbb{N}$. □

Proposition 11.2 (The Darboux property of non-atomic measures)**.** *Let Σ be a σ-algebra and $\mu\colon \Sigma \to [0,\infty)$ be a non-atomic, σ-additive measure. Then for every $A \in \Sigma$ and $t \in [0, \mu(A)]$ there exists a set $B \subset A$, $B \in \Sigma$, such that $\mu(B) = t$.*

Proof. Let $A_0 = A$ and define

$$\alpha_1 = \sup\big\{u \leqslant t\colon u = \mu(B) \text{ for some } B \subset A_0,\, B \in \Sigma\big\}.$$

Pick any set $B_1 \subset A_0$, $B_1 \in \Sigma$, with $\alpha_1 - 1 < \mu(B_1) \leqslant t$. Now, let $A_1 = A_0 \setminus B_1$ and define

$$\alpha_2 = \sup\big\{u \leqslant t - \mu(B_1)\colon u = \mu(B) \text{ for some } B \subset A_1,\, B \in \Sigma\big\}.$$

Pick any set $B_2 \subset A_1$, $B_2 \in \Sigma$, with $\alpha_2 - 1/2 < \mu(B_2) \leqslant t - \mu(B_1)$. Continuing in this way we get a sequence $(B_n)_{n=1}^{\infty}$ of Σ-measurable, mutually disjoint subsets of A which satisfy $\alpha_n - 1/n < \mu(B_n) \leqslant t - \mu(\bigcup_{j=1}^{n-1} B_j)$ for each $n \in \mathbb{N}$, where

$$\alpha_n = \sup\left\{u \leqslant t - \mu\left(\bigcup_{j=1}^{n-1} B_j\right)\colon u = \mu(B) \text{ for some } B \subset A_{n-1},\, B \in \Sigma\right\}$$

and $A_n = A_{n-1} \setminus B_n$, for $n \in \mathbb{N}$.

Now, set $B = \bigcup_{n=1}^{\infty} B_n$. Obviously, $B \subset A$ and $B \in \Sigma$. Moreover, by the choice of B_n's, for every $n \in \mathbb{N}$ we have $\mu(\bigcup_{j=1}^{n} B_j) \leqslant t$, thus $\mu(B) \leqslant t$. Suppose that $\mu(B) < t$. Then, Lemma 11.1 implies that there exists a set $C \subset A \setminus B$, $C \in \Sigma$, satisfying $0 < \mu(C) < t - \mu(B)$. Consequently, for any $n \in \mathbb{N}$ we have

$$\alpha_n \geqslant \mu(C) + \mu(B_n) > \mu(C) + \alpha_n - \frac{1}{n},$$

which cannot hold true for every n. □

For considering multidimensional versions of Proposition 11.2 we need to decide how to define an *atom* for a vector measure. There are two obvious choices which, fortunately, happen to be equivalent.

Lemma 11.3. *Let Σ be a σ-algebra, X be a Banach space and $\mu\colon \Sigma \to X$ be a σ-additive vector measure of bounded variation. Then, for every $A \in \Sigma$ the following two assertions are equivalent:*

- (i) *$\mu(A) \neq 0$ and for every $B \subset A$, $B \in \Sigma$, we have $\mu(B) \in \{0, \mu(A)\}$;*
- (ii) *$|\mu|(A) \neq 0$ and for every $B \subset A$, $B \in \Sigma$, we have $|\mu|(B) \in \{0, |\mu|(A)\}$ (that is, A is an atom for $|\mu|$ in the usual sense).*

Proof. Assume that A satisfies (i). Then, obviously, $|\mu|(A) \neq 0$ and for every partition $\{A_1, \ldots, A_k\} \in \Pi(A)$ exactly one index $j \in [k]$ satisfies $\mu(A_j) = \mu(A)$, while for every $i \in [k]$, $i \neq j$, we have $\mu(A_i) = 0$. Hence, $|\mu|(A) = \|\mu(A)\|$ and for every $B \subset A$, $B \in \Sigma$, we have

$$|\mu|(B) \leqslant |\mu|(A) = \|\mu(A)\|. \tag{11.1}$$

For every partition $\{B_1, \ldots, B_m\} \in \Pi(B)$ and every $j \in [m]$ we have $\mu(B_j) \in \{0, \mu(A)\}$, thus $\sum_{j=1}^{m} \|\mu(B_j)\|$ is equal either to 0 or to $n\|\mu(A)\|$, for some $n \in \mathbb{N}$, whence inequality (11.1) implies that it must be either 0 or $\|\mu(A)\|$. Consequently, $|\mu(B)| \in \{0, |\mu(A)|\}$.

Now, assume that A satisfies (ii). Then, for some partition $\{A_1, \ldots, A_k\} \in \Pi(A)$ we have $\sum_{j=1}^{k} \|\mu(A_j)\| > 0$. If $\mu(A)$ was 0, then for at least two indices $j \in [k]$, we would have $\mu(A_j) \neq 0$. But then for any such j we would get $0 < |\mu|(A_j) < |\mu|(A)$ which contradicts our assumption. Therefore, $\mu(A) \neq 0$. Now, fix any $B \subset A$, $B \in \Sigma$. If $\mu(B) \notin \{0, \mu(A)\}$, then both $\mu(B)$ and $\mu(A \setminus B)$ are non-zero, thus $0 < |\mu|(B) < |\mu|(A)$ (here we use the assumption that $|\mu|$ is bounded) which contradicts clause (ii). □

Example 11.4. The implication (ii) ⇒ (i) of Lemma 11.3 is not true in general for σ-additive measures of unbounded variation. To see this, let Σ be the σ-algebra of all Lebesgue measurable subsets of $[0,1]$, λ be the Lebesgue measure on $[0,1]$ and for any $p \in (1, \infty)$ let $\mu\colon \Sigma \to L_p[0,1]$ be defined by $\mu(A) = \mathbb{1}_A$. Plainly, μ is a vector measure which has no atoms in the sense of clause (i) of Lemma 11.3. Moreover, for any sequence $(A_j)_{j=1}^{\infty}$ of pairwise disjoint sets from Σ we have

$$\left\| \mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) - \sum_{j=1}^{n} \mu(A_j) \right\|^p = \lambda\Big(\bigcup_{j=n+1}^{\infty} A_j\Big) \xrightarrow[n\to\infty]{} 0,$$

whence μ is σ-additive.

Obviously, $|\mu|(A) = 0$ whenever $\lambda(A) = 0$. Now, suppose $A \in \Sigma$ and $\lambda A > 0$, and for an arbitrary $k \in \mathbb{N}$ consider any partition $\{A_1, \ldots, A_k\} \in \Pi(A)$ such that $\lambda(A_j) = \frac{1}{k}\lambda(A)$ for each $j \in [k]$. Then

$$\sum_{j=1}^{k} \|\mu(A_j)\| = \sum_{j=1}^{k} \frac{1}{k^{1/p}} \lambda(A)^{1/p} = k^{1-1/p} \lambda(A)^{1/p} \xrightarrow[k\to\infty]{} 0,$$

thus $|\mu|(A) = \infty$. Consequently, the variation $|\mu|$ has infinitely many atoms in the sense of clause (ii) of Lemma 11.3.

Of course, if X is finite-dimensional (and this case will be of interest for us), then every σ-additive measure defined on a σ-algebra and taking values in X is bounded (as every strongly additive measure, even defined merely on algebras; see Corollary I.1.19 in [DU77]), whence it is of bounded variation because semivariation and variation coincides in the finite-dimensional case.

In light of Lemma 11.3, for any vector measure $\mu\colon \Sigma \to X$ as above, we may define an *atom* to be a Σ-measurable set A satisfying any (and hence both) of the conditions (i) and (ii). The measure μ is called *non-atomic* if it has no atoms.

Lyapunov's theorem is by no means an easy corollary from its one-dimensional version proved in Proposition 11.2. Since 1940, it has been reproved many times, but any attempt requires a heavy machinery or a lot of ingenuity. For instance, Lindenstrauss [Lin66] proved Lyapunov's theorem by combining the Banach–Alaoglu, Radon–Nikodým and Krein–Milman theorems (see [Rud91, Theorem 5.5]). Ross [Ros05] found a very tricky and elementary proof which uses only the Darboux intermediate value theorem.

At this point, let us mention that the assertion of Lyapunov's theorem is not true in general for infinite-dimensional Banach spaces. As a matter of fact, the range of any σ-additive vector measure, defined on a σ-algebra and taking values in any Banach space X, is always relatively weakly compact (in view of the Bartle–Dunford–Schwartz theorem; see [DU77, Corollary I.2.7]), but it may fail to be convex whenever X is infinite-dimensional. That phenomenon is completely explained by the following result due to Knowles [Kno74] (see also [DU77, Theorem IX.1.4]):

Theorem 11.5 (Knowles, 1974)**.** *Let X be a Banach space, Σ be a σ-algebra of subsets of Ω and $m\colon \Sigma \to X$ be a σ-additive vector measure. Then, there exists a σ-additive measure $\mu\colon \Sigma \to [0,\infty)$ such that for each $A \in \Sigma$ the equality $\mu(A) = 0$ is equivalent to $m(A \cap B) = 0$ for all $B \in \Sigma$. Moreover, the following assertions are equivalent:*

(a) *For every $A \in \Sigma$ the set $\{m(A \cap B)\colon B \in \Sigma\}$ is weakly compact and convex.*
(b) *For every $A \in \Sigma$ with $\mu(A) > 0$ the operator $L_\infty(\mu) \ni f \mapsto \int_A f\,\mathrm{d}m$ is not one-to-one on the subspace of $L_\infty(\mu)$ consisting of all functions vanishing outside A.*
(c) *For every $0 \neq f \in L_\infty(\mu)$ there exists $g \in L_\infty(\mu)$ such that $\|fg\|_\infty > 0$, but $\int_\Omega fg\,\mathrm{d}m = 0$.*

In what follows, we will derive an improvement of the (finite-dimensional) Lyapunov convexity theorem which says that the range $\mu(\Sigma)$ of an 'almost' non-atomic, σ-additive measure μ with values in a finite-dimensional normed space is 'almost' midpoint convex. However, we are not able to derive the closedness of that range (which is non-trivial to the same degree as convexity) from Steinitz's lemma. Therefore, the limit case, where 'almost' becomes 'exactly', implies only the convexity of the closure of $\mu(\Sigma)$. Nonetheless, combining that stability result with the fact that $\mu(\Sigma)$ is closed in the finite-dimensional case (see Proposition 11.9 below) gives an insight into the geometric structure of the range of an 'almost' non-atomic vector measure.

To formulate the announced result, we need a piece of notation. Namely, for any vector measure μ as in Lemma 11.3 we define its 'measure of non-atomicity' by

$$\mathsf{at}(\mu) = \sup\big\{\|\mu(A)\|\colon A \text{ is an atom for } \mu \text{ or } A = \varnothing\big\},$$

(we allow $A = \varnothing$ under the 'sup' sign to guarantee that non-atomicity of μ is equivalent to $\mathsf{at}(\mu) = 0$) and for any set $F \subset X$ we define its 'measure of non-convexity' by

$$\mathsf{co}(F) = \sup\Big\{\mathrm{dist}\Big(\frac{x+y}{2}, F\Big)\colon x, y \in F\Big\}$$

(observe that $\mathsf{co}(F) = 0$ means exactly that the closure of F is convex).

Theorem 11.6 (V. Kadets, 1991)**.** *Let X be a finite-dimensional normed space. There exists a constant $k_X \leqslant \dim X$ such that for every σ-additive vector measure $\mu\colon \Sigma \to X$, defined on any σ-algebra Σ, we have*

$$\mathsf{co}(\mu(\Sigma)) \leqslant k_X \cdot \mathsf{at}(\mu).$$

The proof relies on the two following lemmas supported by Steinitz's lemma.

Lemma 11.7. *Let Σ be a σ-algebra, X be a Banach space and $\mu\colon \Sigma \to X$ be a σ-additive vector measure of bounded variation. Then, for every $\alpha > \mathsf{at}(\mu)$ and $A \in \Sigma$ there exists a Σ-measurable partition $\{A_1, \ldots, A_k\}$ of A such that $\|\mu(A_j)\| < \alpha$ for each $j \in [k]$.*

Proof. Since μ is bounded, it has at most countably many atoms. Therefore, we may choose finitely many of them so that the measure of the union of all remaining ones has norm smaller than α. In this way, the proof is reduced to the non-atomic case.

So, let $\Omega \in \Sigma$ be a non-atomic part for μ and let $n \in \mathbb{N}$ be so large that $\frac{1}{n}|\mu|(\Omega) < \varepsilon$. Since $|\mu|$ is bounded and μ is σ-additive, $|\mu|$ is σ-additive as well* (see Proposition I.1.9 in [DU77]). Therefore, an appeal to Proposition 11.2 produces a partition $\{B_1, \ldots, B_n\}$ of Ω such that $|\mu|(B_j) = \frac{1}{n}|\mu|(\Omega)$ for each $j \in [n]$. Hence, $\|\mu(B_j)\| < \varepsilon$ for $j \in [n]$ and the proof is completed. □

Lemma 11.8. *Let $\mu\colon \Sigma \to X$ be as above and suppose that a positive number r has the property that for every $\varepsilon > 0$ and $A \in \Sigma$ there exists a Σ-measurable set $A_\varepsilon \subset A$ such that $\|\mu(A_\varepsilon) - \frac{1}{2}\mu(A)\| \leqslant r + \varepsilon$. Then, $\mathsf{co}(\mu(\Sigma)) \leqslant 2r$.*

Proof. Fix any $x, y \in \mu(\Sigma)$, $x = \mu(A)$ and $y = \mu(B)$ for some $A, B \in \Sigma$. Applying our assumption to the sets $\widetilde{A} := A \setminus (A \cap B)$ and $\widetilde{B} = B \setminus (A \cap B)$ we find some $A_\varepsilon, B_\varepsilon \in \Sigma$ satisfying

$$\Big\|\mu(A_e) - \frac{1}{2}\mu(\widetilde{A})\Big\| \leqslant r + \varepsilon \quad \text{and} \quad \Big\|\mu(B_e) - \frac{1}{2}\mu(\widetilde{B})\Big\| \leqslant r + \varepsilon.$$

Let $z = \mu(A_\varepsilon \cup B_\varepsilon \cup (A \cap B))$. Then,

$$\mathrm{dist}\Big(\frac{x+y}{2}, \mu(\Sigma)\Big) \leqslant \Big\|z - \frac{x+y}{2}\Big\| = \Big\|\mu(A_\varepsilon) + \mu(B_\varepsilon) - \frac{1}{2}\big(\mu(\widetilde{A}) + \mu(\widetilde{B})\big)\Big\| \leqslant 2(r+\varepsilon),$$

which completes the proof. □

Proof of Theorem 11.6. Fix any $A \in \Sigma$ and $\varepsilon > 0$. By Lemma 11.7, there is a Σ-measurable partition $\{A_1, \ldots, A_k\} \in \Pi(A)$ such that $\|\mu(A_j)\| < \mathsf{at}(\mu) + \varepsilon$ for each $j \in [k]$. We wish to apply Lemma 11.8, so we are to find a suitable set A_ε and we will be looking for it among the sets of the form $\bigcup_{j\in J} A_j$, for $J \subset [k]$. For any subset J of $[k]$ denote $A_J = \bigcup_{j\in J} A_j$. Then,

$$\begin{aligned}\min_{J\subset[k]}\Big\|\mu(A_J) - \frac{1}{2}\mu(A)\Big\| &= \min_{J\subset[k]}\Big\|\sum_{j\in J}\mu(A_j) - \frac{1}{2}\sum_{j=1}^{k}\mu(A_j)\Big\| \\ &= \frac{1}{2}\min_{J\subset[k]}\Big\|\sum_{j\in J}\mu(A_j) - \sum_{j\notin J}\mu(A_j)\Big\| = \frac{1}{2}\min_{\varepsilon_j=\pm1}\Big\|\sum_{j=1}^{k}\varepsilon_j\mu(A_j)\Big\|.\end{aligned}$$

Now, by Steinitz's Lemma 10.1 we infer that the last expression is at most equal to $\frac{1}{2}\mathrm{dim}X \cdot \max_j \|\mu(A_j)\|$, thus we may appeal to Lemma 11.8 putting $r = \frac{1}{2}\mathrm{dim}X \cdot \mathsf{at}(\mu)$. □

*This implication is also true without assuming that $|\mu|$ is bounded; in such a case $|\mu|$ may be simply a σ-additive measure with values in $[0, \infty]$. However, as Example 11.4 shows, without assuming that $|\mu|$ is bounded we cannot infer that $|\mu|$ has no atoms from the fact that μ has no atoms in the sense of clause (i) of Lemma 11.3. Consequently, Proposition 11.2 would not be applicable.