November 4, 2022

Lecture 2: Convex localization through hyperplane bisections

Minerva mini-course on Convexity in High dimensions by Bo'az Klartag

In each lecture in this mini-course, except for the first introductory lecture, we explain one technique in convex geometry. We will not always reach the most advanced applications of the techniques, but we will learn the basic ideas, with the hope that students will understand how the method is applied in other situations as well.

Today we discuss a convexity technique whose precursor is Payne and Weinberger '60, going back to Gromov and Milman '87 and to Kannan, Lovász and Simonovits '93. Last week we mentioned two geometric inequalities from the 19^{th}-century. The first is:

Brunn-Minkowski inequality: For any Borel sets $A, B \subseteq \mathbb{R}^n$ and $0 < \lambda < 1$,

$$|(1-\lambda)A + \lambda B| \geq |A|^{1-\lambda}|B|^{\lambda}$$

where here $|\cdot|$ is n-dimensional volume, and $(1-\lambda)A + \lambda B = \{(1-\lambda)x + \lambda y\,;\, x \in A, y \in B\}$. Among closed sets of positive volume, equality holds if and only if A and B are convex and are translates of one another.

There are quite a few proofs of the Brunn-Minkowski inequality: Either by Steiner symmetrization, or by approximating the sets by finite union of boxes and induction on the number of boxes, or by other methods. We will not prove the Brunn-Minkowski inequality now, but in fact most of the techniques that we will study are strong enough to prove this inequality.

Corollary 1. *Let $K \subseteq \mathbb{R}^n$ be a convex body, $E \subseteq \mathbb{R}^n$ a subspace, X a random vector distributed uniformly in K. Then $Proj_E(X)$ is log-concave.*

Thus, even if we are only interested in uniform measures on convex sets, log-concave measures are probably going to show up. We recall that a density f is log-concave if for all $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$,

$$f((1-\lambda)x + \lambda y) \geq f(x)^{1-\lambda} f(y)^{\lambda}.$$

Proof. Write f_E for the density of $Proj_E(X)$, i.e.,

$$f_E(x) = |K \cap (x + E^{\perp})|.$$

By convexity, for any $x, y \in E$,

$$\frac{K \cap (E^{\perp} + x) + K \cap (E^{\perp} + y)}{2} \subseteq K \cap \left(E^{\perp} + \frac{x+y}{2} \right)$$

From the Brunn-Minkowski inequality we thus get that $\sqrt{f_E(x) f_E(y)} \leq f_E((x+y)/2)$. □

Moreover, and somewhat conversely to the corollary, any log-concave probability distribution is approximately a marginal distribution of a uniform distribution on a convex body in high dimension. Log-concavity is a rather stable property:

Proposition 2. *(Prékopa-Leindler, which is a functional version of the Brunn-Minkowski inequality). If X is a log-concave random vector in $\mathbb{R}^n$ and $E \subseteq \mathbb{R}^n$ is a subspace, then also $Proj_E(X)$ is log-concave.*

. . . .

The second 19^{th}-century inequality that we mentioned last week was the Poincaré inequality.

Poincaré inequality (with the Payne-Weinberger constant) Let $K \subseteq \mathbb{R}^n$ be convex, μ a log-concave probability measure on K. Then for any smooth function $f : K \to \mathbb{R}$ with $\int f d\mu = 0$,

$$\int_K f^2 d\mu \leq \frac{Diam^2(K)}{\pi^2} \int_K |\nabla f|^2 d\mu.$$

There is equality in one dimension for $f(x) = \cos x$ in the interval $[0, \pi]$. In n-dimensions, for an almost-equality consider an elongated, thin cylinder, which is nearly a one-dimensional convex set.

This inequality has several proofs, for example it may be proven by using the Bochner formula (see Yong and Zhong '84 or the unified approach in Bakry and Qian '00). We will prove this inequality by convex localization, a technique which reduces certain n-dimensional problems to 1-dimensional ones.

The proof relies on hyperplane bisections. Let K, μ and f be as in the statement of the Poincaré inequality.

Little topological lemma. Fix an $(n-2)$-dimensional affine subspace $E_0 \subseteq \mathbb{R}^n$. Then there exists a hyperplane $H \subseteq \mathbb{R}^n$ containing E_0 such that

$$\int_{K \cap H^+} f d\mu = \int_{K \cap H^-} f d\mu = 0$$

where $H^{\pm}$ are the two half-spaces determined by H.

Proof. For a unit vector $\theta \in E_0^{\perp} \cong \mathbb{R}^2$ consider the hyperplane

$$H(\theta) = E_0 + \mathbb{R}\theta.$$

The map

$$\theta \mapsto \int_{K \cap H(\theta)^+} f d\mu - \int_{K \cap H(\theta)^-} f d\mu$$

is an odd continuous map, and therefore it has to vanish somewhere. (Later on we will replace this proof by a Borsuk-Ulam type argument.) □

Thus we reduced the task of proving a single inequality on the given convex body K, to the task of proving the two inequalities

$$\int_{K\cap H^{\pm}} f^2 d\mu \leq \frac{Diam^2(K\cap H^{\pm})}{\pi^2} \int_{K\cap H^{\pm}} |\nabla f|^2 d\mu$$

on two smaller sets.

Repeat bisecting recursively. After N steps, obtain a partition of K (up to measure zero) into convex sets $K_1, \ldots, K_{2^N}$ with $\int_{K_i} f d\mu = 0$ for all i. Let us show that we can arrange matters so that the pieces of the partition are almost 1-dimensional, and we obtain a decomposition into needles.

Definition 3. *A convex set $P \subseteq \mathbb{R}^n$ is a δ-needle, for $\delta > 0$, if for some line $\ell \subseteq \mathbb{R}^n$,*

$$P \subseteq \ell + \delta \cdot B^n.$$

We assume that $Diam(K) < \infty$ (otherwise, there is nothing to prove).

Proposition 4. *Let $\delta > 0$. Then there exist $N \geq 1$ and a partition of K as above such that K_i is a δ-needle for all i.*

Proof sketch. Pick a dense sequence in the space of affine $(n-2)$-dimensional subspaces in $\mathbb{R}^n$,

$$E_1, E_2, \ldots \subseteq \mathbb{R}^n.$$

At the i^{th}-step, bisect with respect to a hyperplane containing E_i. For sufficiently large N, this works. □

Proof of P-W. We may assume that f and $|\nabla f|$ are uniformly continuous. Fix $\varepsilon > 0$, choose $\delta > 0$ by uniform continuity, and apply the proposition. It suffices to prove that for all i,

$$\int_{K_i} f^2 d\mu \leq (1+\varepsilon)\frac{Diam^2(K_i)}{\pi^2} \int_{K_i} |\nabla f|^2 d\mu.$$

Since $K_i \subseteq \ell_i + \delta \cdot B^n$, we may project to the line ℓ_i. The projection of $1_{K_i}\mu$ is log-concave, by Prékopa-Leindler. Hence it suffices to prove that for any interval $I \subseteq \mathbb{R}$, a log-concave measure ν on I abd a smooth function g on the interval,

$$\int_I g d\nu = 0 \qquad \Longrightarrow \qquad \int_I g^2 d\nu \leq \frac{Length^2(I)}{\pi^2} \int_I (g')^2 d\nu.$$

By scaling we may assume that $I = [0, \pi]$. Such inequalities are studied in the Sturm-Liouville theory in ordinary differential equations. This part of the argument is not immediately related to convex localization. By standard functional analysis, the inequality that we need to prove amounts to a spectral gap for the operator

$$Lg = e^{\psi}(e^{-\psi} g')' = g'' - \psi' g'$$

with Neumann boundary conditions where $\frac{d\nu}{dx} = e^{-\psi}$ so that $\psi : [0,\pi] \to \mathbb{R}$ is convex. The operator L is self-adjoint with discrete spectrum in $L^2(\nu)$ with its domain being all g such that g, g' are absolutely-continuous with Lg square integrable and

$$g'(0) = g'(\pi) = 0.$$

We know that the first eigenvalue of $-L$ equals zero, and we need to prove that the second eigenvalue is at least one. There is a nice conjugacy transform on the subspace of smooth functions orthogonal to constants (it is an isometry from $\dot{H}^1(\nu)$ to $L^2([0,\pi])$). Setting

$$Tu = e^{-\psi/2}u'$$

we have $ATu = TLu$ where A is a Schrödinger operator

$$Aw = e^{-\psi/2}\left(e^{\psi}(e^{-\psi/2}w)'\right)' = w'' - \left(\frac{\psi''}{2} + \frac{(\psi')^2}{4}\right)w.$$

The operator A is self-adjoint in $L^2([0,\pi])$ with Dirichlet boundary conditions. Since $\psi'' \geq 0$, the operator $-A$ is at least, in the operator sense, the operator $-w''$ with Dirichlet boundary conditions. The spectrum of the latter operator is

$$1, 4, 9, 16, \ldots$$

Hence $-A \geq \mathrm{Id}$ and the spectral gap of L is at least one, as any eigenvalue of L (except zero) is also an eigenvalue of A. □

. . . .

The next application of Convex Localization that we would like to present is Gromov's Gaussian waist inequality. It is more complicated that the Payne-Weinberger proof, but its main is rather intuitive. Write γ_n for the standard Gaussian measure in $\mathbb{R}^n$, with density

$$(2\pi)^{-n/2}e^{-|x|^2/2}.$$

Theorem 5 (Gromov '02)**.** *Let $1 \leq k \leq n$ and let $f : \mathbb{R}^n \to \mathbb{R}^k$ be a continuous function. Then there exists $t \in \mathbb{R}^k$ such that the fiber $L = f^{-1}(t)$ satisfies*

$$\gamma_n(L + rB^n) \geq \gamma_n(\mathbb{R}^{n-k} + rB^n) \qquad \textit{for all } r > 0. \tag{1}$$

There is equality when f is linear. The case $k = 1$ follows from the Gaussian isoperimetric inequality, with t being the median of f. Unlike the case $k = 1$ where most of the mass is in large fibers, the case $k \geq 2$ is a bit like finding a "needle in a haystack". It could happen that most of the mass belong to small fibers (Alpert-Guth '14).

We may that μ is "more log-concave than the Gaussian" if $d\mu = e^{-\psi} dx$ and

$$\nabla^2 \psi \geq \mathrm{Id},$$

in the case where ψ is smooth. When ψ is not smooth, we require that $\psi(x) - |x|^2/2$ is convex, where the function ψ is in fact allowed to attain the value $+\infty$. We may replace γ_n by such μ on the left-hand side of (1), this is the version that we will prove. This is a theorem about convexity, not about the symmetries and structure of the Gaussian measure.

Suppose that μ is more log-concave than the standard Gaussian. Then first, the restriction of μ to any open convex set is still more log-concave than the standard Gaussian. Second, the same applies for the orthogonal projection of μ to any subspace, as follows from the Prékopa-Leindler inequality.

There is also a spherical version of the waist inequality (Gromov '02), with a similar proof. There are also versions the cube (K. '17) and for the Euclidean ball (Akopyan and Karasev '17), as well as less precise versions for general convex bodies (K. '17).

We proceed with a proof of Theorem 5 following K '17. Unlike the case of Payne-Weinberger, here the only proof that I know uses convex localization.

The main property of log-concave measures that we use is the following

Lemma 6. *Let ν be a probability measure on $\mathbb{R}^k$ that is more log-concave than the standard Gaussian measure. Then there exists $x_0 \in \mathbb{R}^k$, referred to here as $Center(\nu)$ such that*

$$\nu(x_0 + rB^k) \geq \gamma_k(rB^k) \qquad \textit{for all } r > 0.$$

Proof. Suppose that $d\nu/dx = e^{-\psi}$ with ψ smooth, thus $\nabla^2\psi \geq \mathrm{Id}$. We pick $x_0 = \mathrm{argmin}\psi$, which is uniquely determined, and assume that $x_0 = 0$. Then $\nabla\psi(0) = 0$ and for all $x \in \mathbb{R}^n$,

$$\partial_r \psi(x) \geq |x|.$$

Thus on each ray emanating from the origin, the function $e^{-\psi}$ decays faster than $e^{-|x|^2/2}$ and in fact $e^{-\psi(x)}/e^{-|x|^2/2}$ is decreasing along any ray emanating from the boundary. For any $\theta \in S^{n-1}$, let $\alpha = \alpha(\theta)$ be such that

$$\int_0^\infty [\alpha e^{-\psi(t\theta)} - e^{-t^2/2}] t^{k-1} dt = 0. \tag{2}$$

Then the integrand is non-negative up to some $t_0 \geq 0$, and afterwards it is non-positive. Therefore, if we integrate in (2) from 0 to r, we get a unimodal function of $r \in [0, \infty)$ that vanishes at zero and infinity. Hence it is non-negative. This means that

$$\int_0^r e^{-\psi(t\theta)} t^{k-1} dt \geq \alpha^{-1} \int_0^r e^{-t^2/2} t^{k-1} dt = \gamma_k(rB^k) \cdot \int_0^\infty e^{-\psi(t\theta)} t^{k-1} dt.$$

We now integrate over $\theta \in S^{k-1}$ and use polar coordinates in $\mathbb{R}^k$ to conclude the lemma. □

It may be arranged so that the center of ν varies continuously with ν in the weak* topology (perhaps after allowing a little slack of small $\varepsilon > 0$, we are not entirely accurate here about the definition of the center, it needs to be regularized, see K. 17 for precise definition). Given a continuous function $f : \mathbb{R}^n \to \mathbb{R}^k$ and an open convex set $K \subseteq \mathbb{R}^n$ we define

$$\mathcal{I}(K) = f(\widetilde{Center}(\mu|_K)) \in \mathbb{R}^k,$$

where $\widetilde{Center}$ is a certain regularized variant of $Center$. Fix $\delta > 0$. Our goal is to create a partition up to measure zero

$$\mathbb{R}^N = K_1 \cup \ldots \cup K_{2^N}$$

for some number $N \geq 1$ and convex pieces $K_1, \ldots, K_N \subseteq \mathbb{R}^n$, such that two properties hold:

(i) $\mathcal{I}(K_1) = \ldots = \mathcal{I}(K_{2^N})$. This would follow from a variant of the Borsuk-Ulam theorem, using nothing but the continuity of $\mathcal{I}$.

(ii) Each K_i is a (k, R, δ)-pancake. A convex set $P \subseteq \mathbb{R}^n$ is a (R, k, δ)-pancake if there exists an affine k-dimensional subspace $E \subseteq \mathbb{R}^n$ such that

$$P \cap RB^n \subseteq E + \delta \cdot B^n.$$

Here $R > 0$ is thought of as very large and $\delta > 0$ is very small.

Assuming these two properties for now, we can proceed as follows:

Sketch of proof of Theorem 5. We define the center point slightly differently, so that the following holds for all of our pancakes: Set

$$\nu_i = Proj_{E_i}(\mu|_{K_i}),$$

the push-forward of the restriction of μ to K_i to the subspace E_i with $K_i \cap RB^n \subseteq E_i + \delta \cdot B^n$. Then ν_i is more log-concave than the standard Gaussian. Now, for $x_0 = \widetilde{Center}(\mu|_{K_i})$,

$$\mu((x_0 + rB^k) \cap K_i) \geq \nu_i(Center(\nu_i) + rB^k) \cdot (1 + \varepsilon(\delta))$$

for some small error ε depending on δ, uniformly in r in some interval that approximates $(0, \infty)$.

Set $t = f(\widetilde{Center}(\mu|_{K_i}) \in \mathbb{R}^k$, which is the same for any i. Thus the fiber $L = f^{-1}(t)$ passes through the centers of all convex pieces of the partition. Therefore

$$\mu((L + rB^k) \cap K_i) \geq \nu_i(Center(\nu_i) + rB^k) \cdot (1 + \varepsilon(\delta)) \geq \nu_i(E_i) \cdot \gamma_k(rB^k) \cdot (1 + \varepsilon(\delta)).$$

Hence by summing over i and we obtain

$$\mu(L+rB^k) = \sum_{i=1}^{2^N} \mu((L+rB^k) \cap K_i) \geq \sum_{i=1}^{2^N} \mu(K_i) \cdot \gamma_k(rB^k) + \varepsilon(\delta) = \gamma_n(\mathbb{R}^{n-k} + rB^k) + \varepsilon(\delta),$$

and the theorem follows by letting δ tend to zero. $\square$

How can we prove the two properties above? Let $1 \leq k \leq n, N \geq 1$ and let $E_1, \ldots, E_N \subseteq \mathbb{R}^N$ be any fixed $(N-k-1)$-dimensional affine subspaces.

We apply hyperplane bisection recursively N steps, and at the i^{th}-step we will bisect with respect to hyperplanes containing E_i. The space of all such oriented hyperplanes is an k-dimensional sphere. Thus we have freedom to solve k equations with the choice of each hyperplane.

Theorem 7 (Borsuk-Ulam type). *We can choose such hyperplanes such that the partition of K into $K_1, \ldots, K_{2^N}$ that we obtain satisfies $\mathcal{I}(K_1) = \mathcal{I}(K_2) = \ldots = \mathcal{I}(K_{2^N})$.*

Only the continuity of $\mathcal{I}$ is relevant for this theorem. Note that for $N = 1$ this is exactly the Borsuk-Ulam theorem. There are a few ways to prove Theorem 7. One possibility is to adapt the homotopy proof of Borsuk-Ulam, replacing the $\pm$-symmetry by the group of symmetries of the complete binary tree, which also acts on the space of partitions of K obtained by convex localization. Another possibility is to use the axioms of Stiefel-Whitney classes, and show that any section of a suitable vector bundle has to vanish. Why do we obtain pancakes?

Proposition 8. *Let $0 \leq k \leq n-1$ and $R \geq 1, 0 < \delta < 1$. Then there exist $N \geq 1$ and $(n-k-1)$-dimensional affine subspaces $E_1, \ldots, E_N \subseteq \mathbb{R}^n$ with the following property: If H_i is a hyperplane containing E_i, then for any choice of signs*

$$H_1^{\pm} \cap \ldots \cap H_N^{\pm}$$

is an (R, k, δ)-pancake. Here H^+ and H^- are the two half-spaces whose boundary is H.

This is proven analogously to Proposition 4, by selecting any dense sequence $E_1, E_2, \ldots$ in the space of affine $(n-k-1)$-dimensional subspaces in $\mathbb{R}^n$, and picking a sufficiently large N.

We conclude this lecture by mentioning one more application of convex localization, proven by Bourgain '91, with subsequent improvements by Bobkov '00, Carbery and Wright '01, Nazarov, Sodin and Volberg '03 and Fradelizi '09.

Theorem 9. *(reverse Hölder inequality for polynomials) Let X be a log-concave random vector in $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial whose degree is at most d. Then for $0 \leq p \leq q$,*

$$\|f(X)\|_p \leq \|f(X)\|_q \leq C\|f(X)\|_p$$

where $C = C_{p,q,d}$ is independent of X or n. One can prove $C_{p,q,d} \leq C^d[(q+1)/(p+1)]^d$.

For $d = 1$ this was proven by Berwald '47 using the Brunn-Minkowski inequality. This statement reduces to one dimension (convex localization where f has the same L^p-norm on all pieces). In one dimension, a log-concave distribution has sub-exponential tail, hence the theorem holds true for $d = 1$. In general, we may multiply f by a constant so that $f(x) = \prod_{i=1}^d (x - z_i)$. By Hölder inequality and the case $d = 1$,

$$\left\| \prod_{i=1}^d (X - z_i) \right\|_q \leq \prod_{i=1}^d \|X - z_i\|_{qd} \leq C_{qd} \prod_{i=1}^n \|X - z_i\|_0 = C_{qd}\|f(X)\|_0,$$

as $\|fg\|_0 = \|f\|_0 \cdot \|g\|_0$.