On nearly-radial marginals of high-dimensional probability measures

Bo’az Klartag

Tel Aviv University

Conference on Convex and Discrete Geometry
in honor of Peter M. Gruber

Vienna, July 2009

A convex inspiration

We will discuss a certain theorem which does not seem to belong in the realm of Convex and Discrete Geometry.

- Yet, the spirit and the approach of high-dimensional convex geometry are essential for this result.

Theorem (Dvoretzky '60, Milman '71)

Suppose $K \subset \mathbb{R}^n$ is a convex body, $\ell \leq \lfloor c \log n \rfloor$. Then there exists an ℓ-dimensional subspace $F \subset \mathbb{R}^n$ for which $\pi_F(K)$ is approximately a Euclidean ball.

What exactly is "approximately a Euclidean ball"?

$$B \subseteq \pi_F(K) \subseteq 1.01B$$

where $B \subset F$ is a Euclidean ball centered at the origin.

Duality

No direct proof. By duality, an equivalent formulation is:

Theorem (Dvoretzky '60, Milman '71)

Suppose $K \subset \mathbb{R}^n$ is a convex body, the origin in its interior, $\ell \leq \lfloor c \log n \rfloor$. Then there exists an ℓ-dimensional subspace $F \subset \mathbb{R}^n$ for which $K \cap F$ is approximately a Euclidean ball.

- Idea of Milman's proof: The **concentration of measure.** When $f : S^{n-1} \to \mathbb{R}$ is a Lipschitz function, its restriction to a typical subspace is approximately constant.
- Furthermore, in the appropriate sense, a **random** section of a convex body is usually approximately round.

Measure theoretic formulation

- Rather than projecting a set, we can project a measure.

Definition

Suppose μ is a probability measure on $\mathbb{R}^n$, and $E \subset \mathbb{R}^n$ a subspace. The **marginal** or **measure projection** of μ onto E is the measure $\pi_E(\mu) = (\pi_E)_*(\mu)$ on E for which

$$\pi_E(\mu)(A) = \mu(\pi_E^{-1}(A)) \qquad \forall \text{ Borel } A \subseteq E.$$

Question

What is the measure-theoretic analog of Dvoretzky's theorem?

- Many (but not all) high-dimensional probability measures have approx. gaussian marginals. Examples include: Statistical independence, or thin-shell condition, or convexity assumptions.

Nearly-radial marginals

A remark by Gromov, '88

Perhaps high-dimensional probability measures admit approximately-radial marginals.

... With little extra effort the above discussion applies to projection of measure on $\mathbb{R}^i$ rather than of subsets. Let μ_i be a probability measure on $\mathbb{R}^i$, for all $i = 1, 2, \ldots$, such that the support of μ_i linearly spans $\mathbb{R}^i$.

Then, for every $j = 1, 2, \ldots$, there exists an orthogonally invariant measure $\bar{\mu}$ on $\mathbb{R}^j$ and a sequence of linear maps $A_i : \mathbb{R}^i \to \mathbb{R}^j$, such that the push-forward measures $A_*(\mu_i)$ on $\mathbb{R}^j$ weakly converge to $\bar{\mu}$ ...

- Some technical details need to be fixed, but the idea is certainly correct and interesting.

Explanation of Gromov's remark

We need to explain what is

1. A "high-dimensional" probability measure.
2. An "approximately-radial" measure.

Suppose μ is a probability measure on $\mathbb{R}^d$. When is μ approximately radial?

- We look for a definition which is scale-invariant, and corresponds to weak convergence of probability measures.

Easy case:

Suppose μ is supported on the sphere $S^{d-1} \subset \mathbb{R}^d$, and $\varepsilon > 0$. We will say that μ is ε-radial if

$$W_1(\mu, \sigma_{d-1}) \leq \varepsilon.$$

i.e., the L^1 Monge-Kantarovich distance between μ and the uniform measure on the sphere is small.

Measures on a thin spherical shell

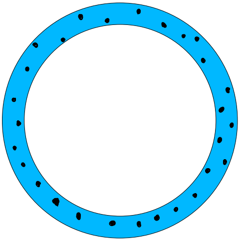
Recall: $$W_1(\mu,\nu) = \sup\left[\int \varphi d\mu - \int \varphi d\nu\right]$$
where the supremum runs over all 1-Lipschitz functions.

- Why did we choose L^1 Monge-Kantorovich metric?

For no particular reason. We may take L^p for $1 \leq p < \infty$, or discrepancy metric. Any reasonable metric that metrizes weak convergence will probably work.

A slightly more general case:

Suppose μ is supported on a *thin spherical shell* $\{a \leq |x| \leq b\}$ with $a \approx b$. We say that μ is ε-radial if $\mathcal{R}_*(\mu)$ is ε-radial as a measure on the sphere. Here, $\mathcal{R}(x) = x/|x|$ is radial projection.



Nearly-radial measures

- For a measure μ and a set A of positive measure, write $\mu|_A$ for the conditioning of μ to A, that is,

$$\mu|_A(B) = \mu(A \cap B)/\mu(A).$$

- A spherical shell is a set $S \subset \mathbb{R}^d$ of the form $S = \{x \in \mathbb{R}^d; a \leq |x| \leq b\}$ with $a, b > 0$.

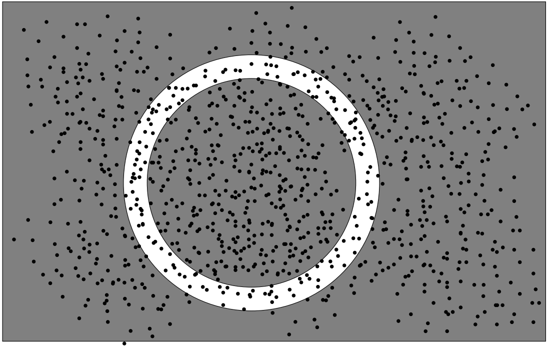
Definition (Gromov)

Let μ be a prob. on $\mathbb{R}^d$ with $\mu(\{0\}) = 0$ and let $\varepsilon > 0$. We say that μ is ε-radial if for any spherical shell $S \subset \mathbb{R}^d$ with $\mu(S) \geq \varepsilon$,

$$W_1\left(\mathcal{R}_*(\mu|_S), \sigma_{d-1}\right) \leq \varepsilon.$$

i.e., when we condition μ to the shell S, and project radially to the unit sphere, the resulting prob. measure is ε-close to the uniform measure on S^{d-1} in the W_1 metric.

Example:



Theorem

Let μ be an absolutely continuous probability measure on $\mathbb{R}^n$, and let $n \geq (C/\varepsilon)^{Cd}$. Then, there exists a linear projection $\pi : \mathbb{R}^n \to \mathbb{R}^d$ with $\pi(\mu)$ an ε-radial measure on $\mathbb{R}^d$.

Remarks

- Thus, any absolutely-continuous measure in $\mathbb{R}^n$ has ε-radial marginals of dimension $d = \lfloor c \log n / \log(1/\varepsilon) \rfloor$.
- Such dependence on ε is only conjectured in the context of Dvoretzky's theorem. Current estimates are weaker.
- Gromov had a topological proof for the case $d = 1$ and $d = 2$. The case $d = 1$ means that the measure is approx. symmetrical on the real line.

Do we have to assume that μ is absolutely continuous?

- Suppose μ is a combination of a gaussian measure and several atoms. None of the marginals are approx. radial.

When is a measure truly high-dimensional?

Definition

A prob. measure μ on $\mathbb{R}^n$ is "decently high-dimensional with accuracy δ", or $1/\delta$-*dimensional* in short, if

$$\mu(E) \leq \delta \dim(E)$$

for any subspace $E \subseteq \mathbb{R}^n$.

- We say that μ is *decent* if it is n-dimensional.

Of course, all absolutely-continuous measures are decent, as well as many discrete measures.

- Examples for decent measures: The uniform measure on the cube's vertices, or on n linearly independent vectors.

Main result

Theorem

Let μ be a decent probability measure on $\mathbb{R}^n$, $\varepsilon > 0$, and let

$$d \leq c\frac{\log n}{\log 1/\varepsilon}.$$

Then, there exists a linear projection $\pi : \mathbb{R}^n \to \mathbb{R}^d$ such that $\pi(\mu)$ is an ε-radial measure on $\mathbb{R}^d$.

When μ is $1/\delta$-dimensional, for $\delta > 0$, we may take $\varepsilon = c\delta^{c/d}$.

- Most marginals of a high-dim. measure are approximately spherically-symmetric, with almost no assumptions.

Criticism

The definition of an ε-radial measure is too complicated. Is it good for anything?

An application to tail distributions

- Any high-dim. measure has super-Gaussian marginals. There are no decent random vectors X in $\mathbb{R}^n$ for which $\varphi(X)$ has a tail like $\exp(-t^3)$ for all linear functionals φ.

Corollary

Let X be a decent random vector in $\mathbb{R}^n$. Then, there exists a non-zero linear functional φ on $\mathbb{R}^n$ with

$$\mathbb{P}\left(\varphi(X) > tM\right) > c\exp(-Ct^2) \quad \textit{for } 0 \leq t \leq R_n,$$

where M is a median of $|\varphi(X)|$, (so $\mathbb{P}(|\varphi(X)| \geq M) \leq 1/2$), and $R_n = c(\log n)^{1/4}$.

Proof ideas

The proof of the theorem consists of three steps:

Finding the right "position"

First, we apply an appropriate linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $\nu = T(\mu)$ satisfies

$$\int_{\mathbb{R}^n} \left(\frac{x}{|x|} \cdot \theta \right)^2 d\nu(x) = \frac{1}{n} \qquad \forall \theta \in S^{n-1}.$$

- Such a linear transformation exists only for decent measures.
- Compare with the more common "isotropic position".
- Our position appears in the analysis of the general Brascamp-Lieb inequalities (Barthe '98, Carlen-Lieb-Loss '04, Carlen-Cordero '09).

Second step of the proof

Once the measure μ is in the right "position", we decompose our measure:

$$\mu = \int \mu_\alpha + E$$

where

1. Each μ_α is a probability measure, uniform on $n^{1/5}$ vectors in $\mathbb{R}^n$ that are approx. orthogonal (but not of the same length).
2. E is a small error (its total mass is at most $n^{-1/5}$).

Similar constructions were used by Bourgain, Lindenstrauss and Milman '86.

- This decomposition is possible because of the high dimensionality: Randomly select $n^{1/5}$ vectors according to μ. Then with high probability, they are approx. orthogonal.

Completing the proof

In the third and final step of the proof, we randomly project each element of our decomposition.

Projecting an ensemble of almost-orthogonal vectors

Suppose μ is a probability measure supported on ℓ almost-orthogonal vectors in $\mathbb{R}^n$. Let $\varepsilon > 0$ and let $\Gamma : \mathbb{R}^n \to \mathbb{R}^d$ be a gaussian random matrix. Assume that

$$d \approx \log n / (\log 1/\varepsilon), \quad \ell \approx n^{1/5}.$$

Then, with high probability, $\Gamma(\mu)$ is an ε-radial measure on $\mathbb{R}^d$.

Thus, a random projection of the measure μ is approx. radial.

- The proof does not rely directly on the concentration of measure / isoperimetric inequalities in high dimension.

Infinite-dimensional formulation

Suppose μ is a Borel probability measure on X, where $X = \mathbb{R}^\infty$, or X is a separable normed space, or a separable Fréchet space, or dual to such a space.

Theorem

Assume that for any finite-dimensional subspace $E \subset X$,

$$\mu(E) = 0.$$

Then for any d and $\varepsilon > 0$, there exists a continuous linear map $T : X \to \mathbb{R}^d$ with $T(\mu)$ an ε-radial measure.

- We also have “arbitrarily long super-gaussian tails”: For any $R > 0$ there exists a 1D marginal that is super-guassian up to R standard deviations.

Thank you.