Concentration of measure in high-dimensional convex sets

Joseph Lehec
Université de Poitiers

The 5th Korea-France Conference in Mathematics
Institut de Mathématique de Bordeaux
July 15, 2025

Isoperimetry

Consider the sphere $\mathbb{S}^{n-1}$ equipped with its normalized Haar measure σ_{n-1} and geodesic distance d.
For a $A \subset \mathbb{S}^{n-1}$ we define the enlargement of A:

$$A_r := \{x \in \mathbb{S}^{n-1} \colon d(x, A) \leq r\},$$

Theorem [P. Lévy (circa 1950)]

For $A \subset \mathbb{S}^{n-1}$, if C is a spherical cap (i.e. geodesic ball) having the same measure as A then

$$\sigma_{n-1}(A_r) \geq \sigma_{n-1}(C_r).$$

In other words among sets of a given measure, spherical caps minimize the measure of the enlargement.

Corollary [V. Milman]

If $\sigma_{n-1}(A) = \frac{1}{2}$ then

$$\sigma_{n-1}(\mathbb{S}^{n-1} \backslash A_r) \leq \exp\left(-c\, n \cdot r^2\right)$$

where $c > 0$ is a universal constant.

Indeed, by isoperimetry

$$\sigma_{n-1}(\mathbb{S}^{n-1} \backslash A_r) \leq \sigma_{n-1}(\mathbb{S}^{n-1} \backslash C_r)$$

where C is a half-sphere. A elementary computation shows that

$$\sigma_n(\mathbb{S}^{n-1} \backslash C_r) \leq \exp(-cn \cdot r^2).$$

This phenomenon is called measure concentration: the enlargement of a set of measure $1/2$ contains all but a tiny proportion of the space.

This can be reformulated in terms of Lipschitz functions.

Theorem

If $f : \mathbb{S}^{n-1} \to \mathbb{R}$ is Lipshitz with constant L and if m_f is a median for f:

$$\sigma_{n-1}(|f - m_f| \geq r) \leq 2 \exp\left(-c\,n \cdot \frac{r^2}{L^2}\right).$$

Indeed because of the Lipschitz property

$$\{f \leq m_f\}_{r/L} \subset \{f \leq m_f + r\}$$

Then apply measure concentration to $\{f \leq m_f\}$.
Remark: one can replace the median by the mean at the cost of worst constant c.
So another formulation of measure concentration is that Lipschitz functions are very close to their mean with high probability.

This concept of measure concentration was introduced by V. Milman in his revolutionary proof of Dvoretzky's theorem.

Theorem [Dvoretzky '61]

If X a Banach space of infinite dimension, for every $\varepsilon > 0$ and every integer k there exists a subspace of E of X of dimension k whose distance to the Euclidean space of the same dimension is at most $1 + \varepsilon$.

Here the distance is that of Banach-Mazur. More explicitely, this means that there exists a Euclidean norm on E denoted $\| \cdot \|_2$ such that

$$\|x\|_2 \leq \|x\|_X \leq (1 + \varepsilon)\|x\|_2, \quad \forall x \in E.$$

In 1971, V. Milman found the following quantitative version:

Theorem [V. Milman '71]

If X is a Banach of dimension n then most subspaces of dimension $k = c(\varepsilon) \log n$ are $1 + \varepsilon$-close to being Euclidean.

- *Most* subspaces in the sense of the Haar measure on the set k dimensional subspaces
- $\log n$ is optimal (for ℓ_∞^n) but certain Banach spaces are better behaved. For instance ℓ_1^n admits Euclidean sections of dimension proportional to n.

Sketch of Milman's proof of Dvoretzky

- Up to a linear transformation, we can assume $\|\cdot\|_X \leq \|\cdot\|_2$ (Eudlidean norm) and

$$M := \int_{\mathbb{S}^{n-1}} \|x\|_X \, d\sigma_{n-1} \geq c \cdot \frac{\sqrt{\log n}}{\sqrt{n}}.$$

- In particular $x \in \mathbb{S}^{n-1} \mapsto \|x\|_X$ is 1-Lipscihtz
- Using concentration of measure we see that for θ uniform on the sphere we have

$$(1-\varepsilon)M \leq \|\theta\|_X \leq (1+\varepsilon)M$$

  with very large probability, namely $\geq 1 - \exp(-c(\varepsilon)nM^2)$.
- By a union bound we can have this property for (say) $\exp(\frac{1}{2}c(\varepsilon)nM^2)$ many points simultaneously and with high probability, which is enough to approximate a subspace of dimension

$$c'(\varepsilon)nM^2 \geq c''(\varepsilon)\log n.$$

Dvoretzky's theorem, especially as proved by Milman, is a milestone in the local (that is, finite-dimensional) theory of Banach spaces. While I feel sorry for a mathematician who cannot see its intrinsic appeal, this appeal on its own does not explain the enormous influence that the proof has had, well beyond Banach space theory, as a result of planting the idea of measure concentration in the minds of many mathematicians. Huge numbers of papers have now been published exploiting this idea or giving new techniques for showing that it holds.

T. Gowers

Let (X, d) be a metric space and let μ be a probability measure on X (equipped with its Borel σ-field)

Definition: Concentration function

$$\alpha_\mu(r) = \sup\left\{\mu(X \backslash A_r) \colon \mu(A) = \frac{1}{2}\right\}$$

Equivalently, α_μ is the smallest function such that

$$\mu(f \geq m_f + r) \leq \alpha_\mu(r)$$

for every 1-Lipschitz f.
We have seen that on $(\mathbb{S}^{n-1}, d, \sigma_{n-1})$

$$\alpha_{\sigma_{n-1}}(r) \leq \exp\left(-cnr^2\right)$$

where c is a universal constant.

Gaussian isoperimetry

Let γ_n be the standard Gaussian measure

$$\gamma_n(dx) = \frac{1}{(2\pi)^{n/2}} \mathrm{e}^{-|x|^2/2}\, dx$$

Theorem [Borell '75] [Sudakov-Tsirel'son '74]

Let $A \subset \mathbb{R}^n$ and $r > 0$:

$$\gamma_n(A_r) \geq \gamma_n(H_r)$$

where H is a halfspace having the same measure as A.

As an immediate corollary we get

$$\alpha_{\gamma_n}(r) = \alpha_{\gamma_1}(r) \leq \exp\left(-\frac{1}{2}r^2\right).$$

Note in particular that the concentration function of γ_n does not depend on n.

This approach to concentration is very limited. Indeed, the uniform measure on the sphere and the Gaussian measure are essentially the only two cases for which the isoperimetric problem can be solved explicitly. For general measures, with less symmetries, it is a hopeless task to solve the isoperimetric problem. So we need other tools to reach concentration.

Here is an important example of a more flexible dimension free concentration result.

Proposition

Let $\mu(dx) = \mathrm{e}^{-V(x)}\, dx$ be a probability measure on $\mathbb{R}^n$, such that the potential V satisfies $\nabla^2 V \geq \rho \cdot \mathrm{Id}$ pointwise for some $\rho > 0$ (i.e. V is uniformly convex) . Then

$$\alpha_\mu(r) \leq 2 \exp\left(-c\rho \cdot r^2\right).$$

- In other words, if the potential of μ is more convex than that of some Gaussian measure then its concentration function is essentially as good as that of this Gaussian.
- Moreover this result admits a soft proof. It indeed follows easily from the Prékopa-Leindler inequality, which is a functional form of the Brunn-Minkowski:

$$|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}.$$

It is natural at this stage to ask about the case $\rho = 0$.
A measure μ on $\mathbb{R}^n$ is called log-concave if

$$d\mu = \mathrm{e}^{-V(x)}\,dx$$

with $V\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convex.

- This is a class of measures that contains uniform measures on convex sets, as well as Gaussian measures.
- This class is also stable under various operations, such as taking products, taking marginals, taking convolution...

What can be said about concentration properties of log-concave measures?

- Gaussian type concentration can't be true in general (think of μ being the exponential measure)
- We should normalize the problem in some way. Let us say that μ is isotropic if

$$\int_{\mathbb{R}^n} x \, \mu(dx) = 0 \quad \text{and} \quad \int_{\mathbb{R}^n} x \otimes x \, \mu(dx) = \mathrm{Id}.$$

Any measure (not supported on a lower dimensional space) admits an affine image that is isotropic.

After normalizing there's no obvious obstruction to having exponential concentration with a dimension free constant:

Question

Could it be the case that if μ is log-concave and isotropic on $\mathbb{R}^n$ then

$$\alpha_\mu(r) \leq 2\exp\left(-\frac{r}{C}\right),$$

where C is a universal constant (not depending on the dimension)?

A classical result of Borell asserts that this holds true if we restrict to halfspaces. So we could reformulate as follows: is it true that in the log-concave case, halfspaces are the worst case for concentration (maybe up to constants)? Note that we have seen that this holds true for Gaussian measures.

This question turns out to be equivalent famous conjecture by Kannan, Lovasz, Simonovits from 1995.

Define the Poincaré constant of μ to be the best constant in the inequality

$$\mathrm{var}_\mu(f) \le C_P(\mu) \int |\nabla f|^2 \, d\mu$$

By a result of E. Milman from 2006, in the log-concave case, exponential concentration and Poincaré are equivalent. So the concentration conjecture from the previous slide is equivalent to:

KLS Conjecture ['95]

Letting

$$\psi_n^2 := \sup\{C_P(\mu) \colon \mu \text{ log-concave and isotropic on } \mathbb{R}^n\}.$$

we have $\psi_n = O(1)$.

Again, an equivalent formulation is to say that in the log-concave case, linear functions should be extremal in Poincaré, up to a constant.

The Kannan, Lovasz, Simonovits conjecture was motivated by questions of algorithmic nature. Indeed, the Poincaré constant of log-concave measures is important for the computational complexity of certain stochastic algorithms, coming from optimization, Bayesian statistic and machine learning.

We won't discuss these applications at all in this talk. Instead we focus on the aftermath of the KLS conjecture in the field of asymtpotic convex geometry.

The philosophy behind the field of asymptotic convex geometry is that in a number of situations some universality phenomena appear in high dimension. We've seen already an instance of this with Dvoretzky's theorem. Here is another example.

Central limit problem for convex sets [Sudakov (early 80s)]

Is it true as the dimension tends to $+\infty$, most marginals a log-concave measure on $\mathbb{R}^n$ are nearly Gaussian?

- For the uniform measure on the ball, marginals have density

$$c_n(1-x^2)^{n/2} \approx c_n \exp\left(-\frac{n}{2}x^2\right)$$

- If X is uniform on the cube $[-1,1]^n$ then $\langle X, \theta \rangle = \sum_{i \leq n} \theta_i X_i$ is a linear combination of i.i.d. variables, henceforth near Gaussian if the ℓ_2-norm of θ is sufficiently spread out over the coordinates.

- In 2003, Anttila, Ball, Perissinaki proved that it is enough to prove the following thin-shell bound to get central limit:

$$\mathbb{P}(||X| - \sqrt{n}| \geq \varepsilon_n \sqrt{n}) \leq \varepsilon_n, \quad \text{for some } \varepsilon_n \to 0,$$

  for every log-concave and isotropic random vector X on $\mathbb{R}^n$.
- Bobkov, Koldobsky ['03] put forward the following conjecture.

Conjecture

$$\sigma_n^2 := \sup \left\{ \mathbb{E}(|X| - \sqrt{n})^2 \right\} = O(1)$$

(again the sup is on all log-concave isotropic X on $\mathbb{R}^n$)

- Implies thin-shell with $\varepsilon_n \approx 1/\sqrt{n}$, which in turn implies some optimal speed of convergence in central limit.
- This is clearly a weaker conjecture than KLS.

Hyperplane (or slicing) conjecture

Conjecture [Bourgain '80]

Let K be a convex set in $\mathbb{R}^n$, of volume 1. Does there exists a hyperplane section of K whose $(n-1)$-dimensional volume is larger than a universal constant?

This sounds like an innocent riddle but it is actually quite deep. We won't have time to elaborate but let us just say that a number of important results from asymptotic convex geometry such as the reversed Brunn-Minkowski inequality of Milman, or the reversed Santaó inequality of Bourgain-Milman become trivialities if slicing were to be true.

- define L_n to be the worst case for convex bodies in $\mathbb{R}^n$ in Bourgain's slicing problem:

$$L_n = \left(\inf_K \sup_H \{ |K \cap H|_{n-1} \} \right)^{-1},$$

  where the infimum is taken on convex sets K of volume 1 and the supremum on affine hyperplanes H.
- Recall that σ_n and ψ_n are the worst thin-shell constant and the worst Poincaré constant respectively. We've seen already that

$$\sigma_n = O(\psi_n)$$

- Moreover it was proved by Eldan and Klartag ['12] that

$$L_n = O(\sigma_n)$$

- In terms of the conjecture we thus have the hierarchy

$$\text{KLS} \Rightarrow \text{Thin-shell} \Rightarrow \text{Slicing}$$

The *original* results were as follows:

- Bourgain ['90] $L_n = O(n^{1/4} \log n)$
- KLS [1995] $\psi_n = O(n^{1/2})$.

For the thin-shell constant nothing was known at the time Bobkov and Koldobski formulated the conjecture
(appart from the bound $\sigma_n = O(\sqrt{n})$ which is trivial).
These problem remained dormant (deemed out of reach) for some time, but over the last 15 years or so there have been a lot of activity around these, and it's fair to say that these conjecture fueled this area or research with many new ideas. Among the tools that were involved in this process we should mention localization techniques, optimal transport, curvature and Bochner formula, semigroup tools, geometric measure theory...

There were a number of progresses on the thin-shell conjecture in the late $2000s$.

- Klartag ['08]: $\sigma_n = o(\sqrt{n})$
  By ABP this is just enough to resolve the central limit problem for convex sets
- Fleury ['09]: refinement along the lines of $\sigma_n = O(n^{1/2-\varepsilon})$.
- Klartag ['09]: Thin-shell conjecture is true under an additional assumption of coordinate symmetries.
- Guédon, E. Milman ['11]: $\sigma_n = O(n^{1/3})$.

More recently, techniques from stochastic calculus made their way to this area and lead to spectacular achievements. It all started with a seminal work of Eldan from 2014, in which he proved that the trivial implication between the KLS conjecture and the weaker thin-shell conjecture can be reversed, at the cost of some logarithm in the dimension.
This is quite remarkable, but more importantly, Eldan came up with a new approach, called stochastic localization, which lead to many subsequent developments.

Eldan's stochastic localization in a nutshell

Let μ be log-concave measure on $\mathbb{R}^n$, let $X \sim \mu$ and (B_s) be a standard Brownian motion independent of X. Let $X_s = X + B_s$ (usual heat flow).

- Time reversal: we set $t = 1/s$
- Conditioning: We let μ_t be the law of X conditioned on X_s.

This defines some sort of interpolation:

- when $t \to 0$, X and X_s become independent so $\mu_t \to \mu$.
- When $t \to \infty$, we have $X_s \to X$, and μ_t converges to a Dirac point mass at X.

The point of this construction is that μ_t is t-uniformly log-concave, so it satisfies a good dimension free concentration property, as we saw earlier. Moreover, one can derive a stochastic differential equation for μ_t (this is related to non linear filtering theory) which allows to control how things evolves along time using Itô's calculus.

- Lee-Vempala ['17] $\psi_n = O(n^{1/4})$
  improves upon $\sigma_n = O(n^{1/3})$ from Guédon-Milman and recovers $L_n = O(n^{1/4})$ from Bourgain.
- Chen's ['21] $\psi_n = n^{o(1)}$. Also breaks down the $n^{1/4}$ of Bourgain, nearly 40 years later. This fantastic breakthrough only comes from a new way of controlling things in stochastic localization.
- Klartag, L. ['22] $\psi_n = O(\log^5 n)$.
  Chen's argument + some tools from spectral theory.
- Jambulapathi, Lee and Vempala ['22] $\psi_n = O(\log^{3.22\cdots} n)$
  (refinement of the later)
- Klartag ['23] $\psi_n = O(\sqrt{\log n})$
  Replaces Chen's method by improved Lichnerowicz inequality.
- Guan ['24] $\sigma_n = O(\log \log n)$
  KL method combined with a delicate bootstrap argument.
- Klartag, L. ['24] $L_n = O(1)$.
  Bourgain's conjecture is finally solved, positively.

What's next?

- Now that slicing is done, is thin-shell under way?
- For KLS though there are some serious obstructions with the current approach.
- The tools developed to solve these problems could be useful in other contexts, such as statistical mechanics and particles systems. As we mentioned already, the recent progress on KLS also have applications in optimization, and in statistics.