Minicourse on information-theoretic geometry of metric measure spaces

Mokshay Madiman

University of Delaware

Topological and Geometrical Structures of Information

Centre International de Rencontres Mathématiques, Marseille

28 August 2017

Outline

- A starting point
  - Isoperimetry and curvature in Geometry
  - Isoperimetry in Probability. Curvature?
  - The Bakry-Émery-Ricci tensor as a curvature notion
- Other approaches to probabilistic curvature
  - Log-concave measures and the Prékopa–Leindler inequality
  - Probabilistic curvature via optimal transport
  - Extension to metric measure spaces (Sturm, Lott–Villani)
- Another appearance of information-theoretic geometry
  - The thin shell property of Gaussians
  - Concentration of information
- Concluding remarks

Isoperimetry in Geometry

Let (M, g) be a (connected, smooth) Riemannian manifold of dimension $n \geq 2$. Let dv be the Riemannian volume element, so that the volume of an open set $A \subset M$ is vol$(A) = \int_A dv$. Say that A is "nice" if we have a meaningful notion of surface area of A, e.g., if

$$\sigma(\partial A) = \lim_{\epsilon \to 0} \frac{\text{vol}(A^\epsilon) - \text{vol}(A)}{\epsilon}$$

exists, where $A^\epsilon = \{x \in M : d(x, y) \leq \epsilon$ for some $y \in A\}$ and d denotes the distance function on M

$Question$: If we consider all "nice" sets of fixed volume c, what are the sets of minimal surface area?

Model examples

- Euclidean space $\mathbb{R}^n$: The *isoperimetric extremizers* are Euclidean balls
- Sphere $\mathbb{S}^n$: The *isoperimetric extremizers* are spherical caps (geodesic balls)

Why important? Central geometric question, with many applications to physics. . . (e.g., shape of soap bubbles)

The role of curvature

The Lévy-Gromov comparison theorem [Gromov '86]

Let (M, g) be a compact Riemannian manifold of dimension $n \geq 2$, and $d\mu = \frac{1}{\text{vol}(M)} dv$ be the normalized volume element. Assume that the Ricci curvature is bounded from below by $k > 0$. If S is the n-sphere with constant Ricci curvature k, and $A \subset M$ is open with $\text{vol}_M(A) = \text{vol}_S(B)$ with B a spherical cap on S, then

$$\sigma_M(\partial A) \geq \sigma_S(\partial B)$$

Recall

- The Ricci curvature tensor is a symmetric bilinear form on the tangent space of M, representing the amount by which the volume of a small wedge of a geodesic ball in a curved Riemannian manifold deviates from that of the standard ball in Euclidean space
- The Ricci curvature is determined by the sectional curvatures of a Riemannian manifold, but generally contains less information
- The model spaces $\mathbb{R}^n$, $\mathbb{S}^n$ and $\mathbb{H}^n$ each have constant sectional curvature; in particular, the Ricci curvature of the n-sphere with radius r is the constant $\frac{n-1}{r^2}$
- We say that Ricci curvature is bounded from below by k (or Ric $\geq k$) if Ric$(\xi, \xi) \geq k$ for every $\xi \in T_pM$ and every $p \in M$

The importance of curvature lower bounds

Multiple comparison theorems for nonnegative Ricci curvature

- Lévy-Gromov comparison theorem: applies for Ricci curvature bounded from below by positive constant
- Bonnet-Myers comparison theorem: If M is (geodesically) complete and Ric $\geq k > 0$, then its diameter is at most $\pi\sqrt{\frac{n-1}{k}}$ (which is the diameter of a sphere of constant Ricci curvature k)
- Bishop-Gromov inequality: If M is complete and Ric $\geq k \in \mathbb{R}$, then for any $r > 0$, the volume of a geodesic ball of radius r in M is at most the volume of a geodesic ball of radius r in the corresponding model space (of constant sectional curvature, i.e., rescaled version of $\mathbb{R}^n$, $\mathbb{S}^n$ or $\mathbb{H}^n$)

On the other hand...

[Lohkamp '94] showed that if $n \geq 3$, then *any* n-manifold can be equipped with a complete Riemannian metric that has negative Ricci curvature. Thus Ricci curvature upper bounds can have no topological implications when $n \geq 3$, which means one cannot have (for example) a volume comparison theorem

Isoperimetry in Probability

Let μ be a probability measure on $\mathbb{R}^n$. For $A \subset \mathbb{R}^n$ open, we say A is "nice" if the *surface area of A with respect to μ*, defined as

$$\mu_+(\partial A) = \lim_{\epsilon \to 0} \frac{\mu(A^\epsilon) - \mu(A)}{\epsilon},$$

exists, where $A^\epsilon = \{x \in \mathbb{R}^n : \|x - y\| \leq \epsilon \text{ for some } y \in A\}$

Question: If we consider all "nice" sets A with $\mu(A) = c$, what are the sets of minimal surface area w.r.t. μ?

Model example

Let γ be the standard Gaussian measure $N(0, I_n)$ on $\mathbb{R}^n$, i.e., with probability density function

$$\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$$

Then the isoperimetric extremizers are half-spaces [Sudakov–Tsirelson '74, Borell '75]

Why important? Enormous implications in probability (also statistics, physics, computer science), especially via the concentration phenomenon...

A probabilistic comparison theorem

Consider the Gaussian density $\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$, and observe that $\log\phi(x)$ is just a quadratic function and hence has constant Hessian: in fact,

$$\mathrm{Hess}(-\log\phi(x)) = \mathrm{Hess}\left[\frac{|x|^2}{2} + \frac{n}{2}\log(2\pi)\right] = I_n$$

The Bakry–Ledoux comparison theorem [Bakry–Ledoux '96]

Let μ be a probability measure on $\mathbb{R}^n$ with density function of the form e^{-U}, where U is C^2 and Hess$(U) \geq I_n$. If A is a nice set with $\mu(A) = \gamma(B)$, with B being a half-space in $\mathbb{R}^n$, then

$$\mu_+(\partial A) \geq \gamma_+(\partial B)$$

Remarks

- This resembles the Lévy-Gromov comparison theorem, with the Gaussian replacing the sphere and Hess(U) replacing the Ricci curvature
- Having any constant positive lower bound on Hess(U) works if we take the Gaussian of corresponding variance

Is there a probabilistic notion of curvature?

It seems that a lower bound on Hess(U) for a probability density function e^{-U} on $\mathbb{R}^n$ (in Probability) has similar effects to a lower bound on Ricci curvature for Riemannian manifolds (in Geometry)... Can a theory of curvature be built in probability?

The Bakry–Émery framework [Bakry–Émery '85]

Builds a probabilistic theory of curvature lower bounds, BUT it applies not directly to measure spaces but to Markov semigroups defined on them in a nontrivial way.

- Motivated by the example of Brownian motion on a manifold (or more generally stochastic diffusions), they consider a class of semigroups of operators that can be specified by a generator and that have a probabilistic interpretation as conditional expectations under a Markov process
- For a connected Riemannian manifold (M, g) equipped with the measure $e^{-U}dv$, they can construct an associated Markov semigroup whose curvature is bounded below (in the Bakry-Émery sense) by k if

$$\text{Hess}(U) + \text{Ric} \geq kI_n$$

- This object is now often called the Bakry-Émery-Ricci tensor, and there exist several comparison theorems for it providing generalizations of those in classical Riemannian geometry [Lott '03, Bakry–Qian '05, Wei–Wylie '09, ...]

Log-concavity

A probability density function f on $\mathbb{R}^n$ is *log-concave* if

$$f(\alpha x + (1-\alpha)y) \geq f(x)^\alpha f(y)^{1-\alpha},$$

for each $x, y \in \mathbb{R}^n$ and each $0 \leq \alpha \leq 1$

Remarks

- A variety of densities is log-concave, including the uniform density on any compact, convex set, the (one-sided or two-sided) exponential density, and any Gaussian
- Deeply studied in probability, statistics, optimization and geometry
- The intuition is that log-concave densities resemble Gaussian densities, e.g., several functional inequalities (Poincaré, logarithmic Sobolev) that hold for Gaussians also hold for appropriate subclasses of log-concave distributions
- Not surprising in hindsight: log-concave measures are precisely those that give $\mathbb{R}^n$ nonnegative curvature!

mile-marker

- A starting point
  - Isoperimetry and curvature in Geometry
  - Isoperimetry in Probability. Curvature?
  - The Bakry-Émery-Ricci tensor as a curvature notion
- **Other approaches to probabilistic curvature**
  - Log-concave measures and the Prékopa–Leindler inequality
  - Probabilistic curvature via optimal transport
  - Extension to metric measure spaces (Sturm, Lott–Villani)
- Another appearance of information-theoretic geometry
  - The thin shell property of Gaussians
  - Concentration of information
- Concluding remarks

Isoperimetry in $\mathbb{R}^n$: Brunn-Minkowski inequality

The most basic measure on $\mathbb{R}^n$ is the Lebesgue measure vol_n. One way to understand isoperimetry here is via the Brunn-Minkowski inequality

Brunn-Minkowski inequality

For Borel sets A, B in $\mathbb{R}^n$,

$$\mathrm{vol}_n(\lambda A + (1-\lambda)B) \geq \mathrm{vol}_n(A)^{\lambda}\mathrm{vol}_n(B)^{1-\lambda}$$

where $\lambda A + (1-\lambda)B = \{\lambda x + (1-\lambda)y : x \in A, y \in B\}$

Remarks

- To see the isoperimetric inequality as a consequence, simply take B to be a Euclidean ball of vanishing radius in the equivalent form $\mathrm{vol}_n(A+B)^{1/n} \geq \mathrm{vol}_n(A)^{1/n} + \mathrm{vol}_n(B)^{1/n} \ldots$
- The Brunn-Minkowski inequality motivates the definition of a *log-concave measure*; the measure μ is said to be log-concave if

$$\mu(\alpha A + (1-\alpha)B) \geq \mu(A)^{\alpha}\mu(B)^{1-\alpha}$$

  for any Borel sets $A, B \subset \mathbb{R}^n$ and each $0 \leq \alpha \leq 1$
- An absolutely continuous (w.r.t Lebesgue measure) probability measure μ on $\mathbb{R}^n$ is log-concave iff it has a log-concave density [Prékopa '73, Borell '74]

The Prékopa–Leindler inequality

If $f, g, h : \mathbb{R}^n \to [0, \infty)$ are integrable functions satisfying, for a given $\lambda \in (0,1)$,

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

for every $x, y \in \mathbb{R}^n$, then

$$\int h \geq \left(\int f\right)^{\lambda} \left(\int g\right)^{1-\lambda}$$

Remarks

- Observe that $f = 1_A, g = 1_B$ and $h = 1_{\lambda A + (1-\lambda)B}$ satisfy the hypothesis; in this case, the conclusion is precisely the Brunn-Minkowski inequality; so the Prékopa–Leindler inequality is often called the functional version of the Brunn-Minkowski inequality
- It can be seen as a reversal of Hölder's inequality, which can be written as

$$\left(\int f\right)^{\lambda} \left(\int g\right)^{1-\lambda} \geq \int f^{\lambda} g^{1-\lambda};$$

  the upper bound is in terms of the integral of

$$h^*(z) = \operatorname{ess\,sup}\{f^{\lambda}(x) g^{1-\lambda}(y) : \lambda x + (1-\lambda)y = z\}$$

The magic of the Prékopa-Leindler inequality

If $f, g, h : \mathbb{R}^n \to [0, \infty)$ are integrable functions satisfying, for a given $\lambda \in (0, 1)$,

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

for every $x, y \in \mathbb{R}^n$, then

$$\int h \geq \left(\int f\right)^{\lambda} \left(\int g\right)^{1-\lambda}$$

Implications

- If f, g, h satisfy the hypothesis, then so do fe^{-U}, ge^{-U} and he^{-U}, when U convex. Thus If $f, g, h : \mathbb{R}^n \to [0, \infty)$ are integrable functions satisfying, for a given $\lambda \in (0, 1)$,

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

  for every $x, y \in \mathbb{R}^n$, then

$$\int h d\mu \geq \left(\int f d\mu\right)^{\lambda} \left(\int g d\mu\right)^{1-\lambda}$$

  when μ has a log-concave density
- In particular, any measure with log-concave density is log-concave

Curvature via the Prékopa-Leindler inequality

Let μ be a probability measure on $\mathbb{R}^n$ with a density of form e^{-V}, where Hess$(V) \geq c \in \mathbb{R}$. Suppose $f, g, h : \mathbb{R}^n \to [0, \infty)$ are integrable functions satisfying, for a given $\lambda \in (0, 1)$,

$$h(\lambda x + (1-\lambda)y) \geq \exp\left\{ - c\lambda(1-\lambda)\frac{|x-y|^2}{2} \right\} f^\lambda(x) g^{1-\lambda}(y)$$

for every $x, y \in \mathbb{R}^d$, then

$$\int h d\mu \geq \left(\int f d\mu \right)^\lambda \left(\int g d\mu \right)^{1-\lambda}.$$

Remarks

- Interpret geometrically for sets
- Implies many consequences of positive curvature such as log-Sobolev inequalities etc. [Bobkov-Ledoux '00]
- [Cordero-McCann-Schmuckenschlager '06] extended this to weighted Riemannian manifolds under lower bound on Bakry-Émery-Ricci tensor

Relative Entropy

Suppose M is a space equipped with a reference measure ℓ. Suppose probability measures μ and ν have densities f and g with respect to ℓ. Then the relative entropy between μ and ν is defined by

$$D(\mu\|\nu) = D(f\|g) = \int f(x) \log \frac{f(x)}{g(x)} d\ell(x)$$

For any μ, ν, $D(\mu\|\nu) \geq 0$, with equality iff $\mu = \nu$

Why is it relevant?

- Relative Entropy is a very useful notion of "distance" between probability measures (non-negative, and dominates several of the usual distances, although non-symmetric)
- Specifically one has Pinsker's inequality: if $d_{TV}(\mu, \nu) := \|f - g\|_{L^1(\ell)}$ is the total variation distance, then $d_{TV}(\mu, \nu)^2 \leq 2D(\mu\|\nu)$

Wasserstein distance

Suppose (M, d) is a metric space, and we have probability measures μ and ν on M. Let $\mathcal{C}(\mu, \nu)$ be the set of all couplings of μ and ν, i.e., the set of all probability measures on $M \times M$ with first marginal μ and second marginal ν. The Wasserstein distance between μ and ν is defined by

$$W_2(\mu, \nu)^2 = \inf_{\pi \in \mathcal{C}(\mu,\nu)} \int_{M \times M} d(x, y)^2 d\pi(x, y)$$

Remarks

- This is a metric on the space of probability measures with finite second moment
- Can be thought of as the minimal cost of transporting material from a configuration distributed according to μ to one distributed according to ν
- Remarkable results of Brenier and McCann imply that if μ and ν are "nice" probability measures on $\mathbb{R}^n$, then the optimal coupling has a particularly nice description
- Allows for an infinitesimal description of Wasserstein distance, and an interpretation of it as the length of the geodesic between μ and ν in the space of probability measures. In particular, one has the *displacement interpolation* μ_s with $s \in [0, 1]$ such that $\mu_0 = \mu$ and $\mu_1 = \nu$

Displacement convexity

Let μ_s be the displacement interpolation between μ and ν. A functional E on the space of probability measures is said to be K-displacement convex if

$$(1-s)E(\mu_0) + sE(\mu_1) - E(\mu_s) \geq \frac{K}{2} s(1-s) W_2^2(\mu_0, \mu_1)$$

Remarks

- In particular, E is 0-displacement convex if $E(\mu_s)$ is convex in $s \in [0, 1]$ along each displacement interpolation path

Geometry of metric spaces equipped with measures

Basic idea

- For a Riemannian manifold with Riemannian volume m,

  Ric$(M) > K \quad \iff \quad D(\cdot\|m)$ is K-displacement-convex on $\mathcal{P}^2(M)$

  [Otto-Villani '00, Cordero-McCann-Schmuckenchlager '01-'06, von Renesse-Sturm '05]

- The entropy condition above only uses the metric structure. Hence, for any metric measure space, *say* that it has "Ricci curvature bounded from below" if this convexity of entropy holds. Is this a sensible definition?

- Answer: Yes [Sturm '05-'06, Lott-Villani '09]

The work of Sturm and Lott–Villani

- Extends the notion of curvature lower bound via displacement convexity of relative entropy to a general class of metric measure spaces
- Proves remarkable closure properties under Gromov-Hausdorff convergence of the set of metric measure spaces with a given curvature lower bound
- Finally provides a synthetic notion of curvature for metric measure spaces without demanding additional infrastructure such as Markov diffusion semigroups on the space
- In fact, they provide a more general curvature-dimension criterion $CD(K, N)$ for metric measures; these hold for Riemannian manifolds iff the manifold has dimension at most N and curvature at least K
- More recently, a number of works [Ambrosio–Gigli–Savare '14–'15, Erbar–Kuwada–Sturm '15, . . .] have noted that one can actually prove equivalence between Bakry-Émery-type curvature conditions and Lott–Sturm–Villani-type curvature conditions

mile-marker

- A starting point
  - Isoperimetry and curvature in Geometry
  - Isoperimetry in Probability. Curvature?
  - The Bakry-Émery-Ricci tensor as a curvature notion
- Other approaches to probabilistic curvature
  - Log-concave measures and the Prékopa–Leindler inequality
  - Probabilistic curvature via optimal transport
  - Extension to metric measure spaces (Sturm, Lott–Villani)
- **Another appearance of information-theoretic geometry**
  - The thin shell property of Gaussians
  - Concentration of information
- Concluding remarks

High-dimensional Gaussians lie in a thin shell

Let $Z \sim N(0, I_n)$, i.e., its density on $\mathbb{R}^n$ is

$$\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$$

When the dimension n is large, the distribution of Z is highly concentrated around the sphere of radius $\sqrt{n}$: note $\mathbf{E}|Z|^2 = \sum_{i=1}^n \mathbf{E} Z_i^2 = n$ and

$$\mathrm{Var}(|Z|^2) = \mathrm{Var}\left(\sum_{i=1}^n Z_i^2\right) = \sum_{i=1}^n \mathrm{Var}(Z_i^2) = 2n$$

so that the standard deviation of $|Z|^2$ is $\sqrt{2n}$, much smaller than the mean n of $|Z|^2$

Can also express this concentration property is through a deviation inequality (Chernoff bound, since $|Z|^2/n$ is just the empirical mean of i.i.d. random variables):

$$\mathbf{P}\left\{\frac{|Z|^2}{n} - 1 > t\right\} \leq \exp\left\{-\frac{n}{2}[t - \log(1+t)]\right\}$$

for the upper tail, and a corresponding upper bound on the lower tail

Thin shells beyond Gaussians?

Like many other facts about Gaussian measures, does this concentration property extend to log-concave measures?

Two possible formulations

- We say that X is *isotropic* if Cov$(X) = I_n$ (this ensures the normalization $\mathbf{E}|X|^2 = n$)

  Q: Is there a universal constant C such that $\mathrm{Var}(|X|^2) \leq Cn$ for every isotropic, log-concave X?

  "Thin shell conjecture" in convex geometry, related to Bourgain's hyperplane conjecture and KLS conjecture. Best known bound due to [Lee-Vempala '16] is $\mathrm{Var}(|X|^2) \leq Cn^{3/2}$

- Since $-\log\phi(x) = \frac{n}{2}\log(2\pi) + \frac{|x|^2}{2}$, the quantity that concentrates in Gaussian case, namely $|Z|^2$, is essentially the logarithm of the Gaussian density function

  Q: Is there a universal constant C such that

  $$\mathrm{Var}(-\log f(X)) \leq Cn$$

  for every log-concave X on $\mathbb{R}^n$?

Entropy and Information Content

Let X be a random vector in $\mathbb{R}^n$, with density f. The random variable

$$\widetilde{h}(X) = -\log f(X)$$

may be thought of as the *information content* of X

Discrete case: $\widetilde{h}(X)$ is the number of bits needed to represent X by an optimal coding scheme [Shannon '48]

Continuous case: No coding interpretation, but may think of it as the log likelihood function in a nonparametric statistical model

The entropy of X is defined by

$$h(X) = -\int f(x) \log f(x)\, dx = \mathbf{E}\widetilde{h}(X)$$

Remarks

- Usual abuse of notation: we write $h(X)$ even though the entropy is a functional depending only on the distribution of X
- $h(X)$ takes values in the extended real line $[-\infty, +\infty]$ (if it exists)

Optimal Varentropy bound for log-concave measures

The *varentropy* of a random vector X is defined as $V(X) = \mathrm{Var}(\widetilde{h}(X))$

Theorem: Given a random vector X in $\mathbb{R}^n$ with log-concave density f,

$$V(X) \le n$$

Remarks

- The bound *does not depend* on f– it is universal over the class of log-concave densities
- The bound is *sharp*: if $X \sim f(x) = e^{-\sum_{i=1}^n x_i} \mathbf{1}_{\{x_1,\ldots,x_n \ge 0\}}$,

  then $V(X) = \mathrm{Var}[\sum_{i=1}^n X_i] = n$

  In fact, $V(X) = n$ also holds for $f(x) \propto e^{-\|x\|_K}$ for *any* norm $\|\cdot\|_K$
- The distribution of $\widetilde{h}(X) - h(X)$ is invariant under any affine transformation of $\mathbb{R}^n$; hence the varentropy $V(X)$ is also affine-invariant
- Discovered independently by [Nguyen '13] and [Wang '14] in their Ph.D. theses, improving [Bobkov–M.'11]. Simplest proof due to [Fradelizi–M.–Wang '16]; another given by [Bolley–Gentil–Guillin '15]

Concentration of $\widetilde{h}(X)$

Theorem 2: [Fradelizi–M.–Wang '16] If X has log-concave density f on $\mathbb{R}^n$, then for any $t > 0$,

$$\mathbf{P}\{\widetilde{h}(X) - h(X) \geq nt\} \leq e^{-nr(t)}$$
$$\mathbf{P}\{\widetilde{h}(X) - h(X) \leq -nt\} \leq e^{-nr(-t)}$$

where

$$r(u) = \begin{cases} u - \log(1+u) & \text{for } -1 < u < \infty \\ +\infty & \text{for } u \leq -1 \end{cases}$$

Remarks

- The probability bound *does not depend* on f– it is universal over the class of log-concave densities
- Note that the function r is convex on $\mathbb{R}$ and has a quadratic behavior in the neighborhood of 0 ($r(u) \sim_0 \frac{u^2}{2}$) and a linear behavior at $+\infty$ ($r(u) \sim_\infty u$)
- The moment generation function bound is sharp, and a product of exponential distributions is again extremal
- Improves significantly the results of [Bobkov–M.'11], who first showed concentration of information in log-concave setting

Typical Sets

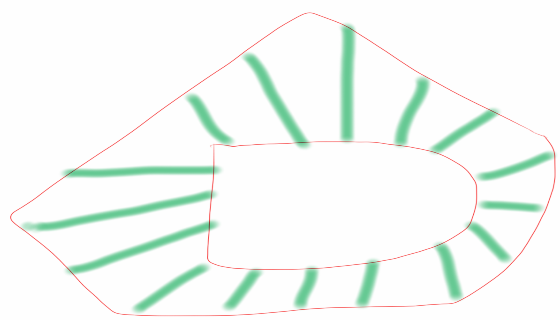
- The Theorem says that for a random vector X in $\mathbb{R}^n$ with log-concave density f, $\frac{1}{n}\log f(X)$ is highly concentrated about its mean

- For fixed $\epsilon > 0$, define the "typical set" by
$$A_\varepsilon = \{x \in \mathbb{R}^n : e^{-h(X)-n\varepsilon} \leq f(x) \leq e^{-h(X)+n\varepsilon}\},$$
i.e., this is the region where $\frac{1}{n}\log f(X)$ deviates from its mean by not more than ϵ

- The typical set can be thought of as the *effective support* of the distribution of X, since X lies in it with high probability

- Moreover, the distribution of X itself is effectively the uniform distribution on the typical set (since $f(X)$ is trapped in a certain range on it)

Typical Sets: A Picture



X is effectively uniformly distributed on the typical set, which is the annulus between two nested convex sets

[Why? Consider the "one-sided" typical set

$$\{x \in \mathbb{R}^n : f(x) \geq C\} \;=\; \{x \in \mathbb{R}^n : -\log f(x) \leq -\log C\}$$

where $C = e^{-h(X)-n\varepsilon}$, as an effective support. This is a convex set.]

Well known example: Standard Gaussian lives effectively in a thin shell around the Euclidean sphere of radius $\sqrt{n}$

Summary

- Notions of curvature are closely related to isoperimetry and concentration results in both Geometry and Probability
- While one framework for curvature lower bounds in the setting of Markov semigroups was provided by Bakry and Émery, more recently a powerful framework has emerged for metric measure spaces due to Lott–Sturm–Villani
- The Lott–Sturm–Villani framework relies on a convexity property of the relative entropy functional
- Entropy also plays other roles in understanding the geometry of probability measures

Thank you for your attention!

○ — ○ — ○