The Kannan-Lovász-Simonovits Conjecture*

Yin Tat Lee[†] Santosh S. Vempala[‡]

December 29, 2017

Abstract

The Kannan-Lovász-Simonovits conjecture says that the Cheeger constant of any logconcave density is achieved to within a universal, dimension-independent constant factor by a hyperplane-induced subset. Here we survey the origin and consequences of the conjecture (in geometry, probability, information theory and algorithms) as well as recent progress resulting in the current best bounds. The conjecture has lead to several techniques of general interest.

Contents

1 Introduction **2**
- 1.1 The KLS conjecture . . . 2
- 1.2 Concentration of measure . . . 3
- 1.3 The original motivation: A sampling algorithm . . . 4
  - 1.3.1 Model of computation . . . 4
  - 1.3.2 Sampling with a Markov chain . . . 4

2 Connections **6**
- 2.1 Geometry and Probability . . . 6
  - 2.1.1 Slicing to anti-concentration . . . 6
  - 2.1.2 Thin shell to central limit theorem . . . 7
  - 2.1.3 Isoperimetry to concentration . . . 8
- 2.2 Algorithms . . . 9
  - 2.2.1 Sampling . . . 9
  - 2.2.2 Sampling to Convex Optimization . . . 9
  - 2.2.3 Sampling to Volume computation and Integration . . . 11

3 Proof techniques **12**
- 3.1 Localization . . . 13
- 3.2 Stochastic localization . . . 16
  - 3.2.1 A stochastic process and its properties . . . 17
  - 3.2.2 Alternative definition of the process . . . 18
  - 3.2.3 Applications . . . 19
- 3.3 Needle decompositions . . . 22

4 Open Problems **23**

---

*This is a **working draft**. We will be grateful for any comments!

[†]University of Washington, yintat@uw.edu

[‡]Georgia Tech, vempala@gatech.edu

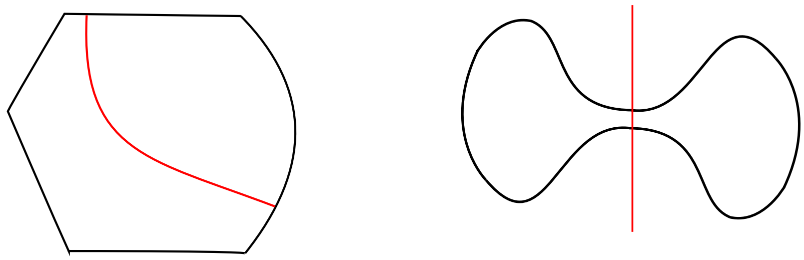


Figure 1.1: Good and bad isoperimetry

1 Introduction

In this article, we describe the origin and consequences of the Kannan-Lovász-Simonovits (KLS) conjecture, which now plays a central role in convex geometry, unifying or implying older conjectures. Progress on the conjecture has lead to new proof techniques and influenced diverse fields including asymptotic convex geometry, functional analysis, probability, information theory, optimization and the theory of algorithms.

1.1 The KLS conjecture

The isoperimetric problem asks for the unit volume set with minimum surface area. For Euclidean space, ancient Greeks (around 150 BC [8]) knew that the solution is a ball; a proof was only found in 1838 by Jakob Steiner [85]. For sets of arbitrary volume, the isoperimetry (or expansion) of the set is defined to be the ratio of surface area to its volume (or its complement, whichever is smaller). For the Gaussian distribution, or the uniform distribution over a Euclidean sphere, the minimum isoperimetric ratio is achieved by a halfspace, i.e., a hyperplane cut [86, 20]. However, this is not true in general (even for the uniform distribution over a simplex) and the minimum ratio set, in the worst case, can be very far from a hyperplane. In general, any domain that can be viewed as two large parts with a small boundary between them, a "dumbbell"-like shape, can have arbitrary isoperimetric ratio (Figure 1.1). It is natural to expect that convex bodies and logconcave functions (whose logarithms are concave along every line) have good isoperimetry — they cannot look like dumbbells. The KLS conjecture says that a hyperplane cut achieves the minimum ratio up to a constant factor for the uniform distribution on any convex set, and more generally for any distribution on $\mathbb{R}^n$ with a logconcave density. The constant is universal and independent of the dimension.

Formally, for a density p in $\mathbb{R}^n$, the measure of a set $S \subseteq \mathbb{R}^n$ is $p(S) = \int_S p(x)\,dx$. The boundary measure of this subset is

$$p(\partial S) = \inf_{\varepsilon \to 0^+} \frac{p(\{x : d(x,S) \le \varepsilon\}) - p(S)}{\varepsilon}$$

where $d(x,S)$ is the minimum Euclidean distance between x and S. The isoperimetric constant of p (or Cheeger constant of p) is the minimum possible ratio between the boundary measure of a subset and the measure of the subset among all subsets of measure at most half:

$$\psi_p = \inf_{S \subseteq \mathbb{R}^n} \frac{p(\partial S)}{\min\{p(S), p(\mathbb{R}^n \setminus S)\}}.$$

For a Gaussian distribution and an unit hypercube in $\mathbb{R}^n$, this ratio is a constant independent of the dimension, with the minimum achieved by a halfspace as mentioned. Both of them belong to a much more general class of probability distributions, called logconcave distribution. A probability density function is logconcave if its logarithm is concave along every line, i.e., for any $x, y \in \mathbb{R}^n$ and any $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \ge f(x)^{\lambda} f(y)^{1-\lambda}. \tag{1.1}$$

Many common probability distributions are logconcave e.g., Gaussian, exponential, logistic and gamma distributions. This also includes indicator functions of convex sets, sets with the property that for any two points $x, y \in K$, the line segment $[x,y] \subseteq K$.

In the course of their study of algorithms for computing the volume, in 1995, Kannan, Lovász and Simonovits made the following conjecture.

Conjecture (KLS Conjecture [46])**.** *For any logconcave density p in $\mathbb{R}^n$,*

$$\psi_p \geq c \cdot \inf_{\text{halfspace } H} \frac{p(\partial H)}{\min\{p(H), p(\mathbb{R}^n \setminus H)\}}. \tag{1.2}$$

where c is an absolute, universal constant independent of the dimension and the density p.

For any halfspace H, its expansion is a one-dimensional quantity, namely the expansion of the corresponding interval when the density projected along the normal to the halfspace. Since projections of logconcave densities are also logconcave (Lemma 2), it is not hard to argue that the isoperimetric ratio is $\Theta(1/\sigma_f)$ for any one-dimensional logconcave density with variance σ_f^2. This gives an explicit formula for the right hand side of (1.2). Going forward, we use the notation $a \gtrsim b$ to denote $a \geq c \cdot b$ for some universal constant c independent of the dimension and all parameters under consideration.

Lemma 1. *For any n-dimensional logconcave density with covariance matrix A,*

$$\inf_{\text{halfspace } H} \frac{p(\partial H)}{\min\{p(H), p(\mathbb{R}^n \setminus H)\}} \gtrsim \frac{1}{\sqrt{\|A\|_{\text{op}}}}$$

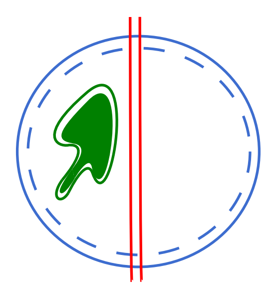
where $\|A\|_{\text{op}}$ is the largest eigenvalue of A, or equivalently, the spectral norm of A.

It is therefore useful to consider the following normalization to a given distribution: apply an affine transformation so that for the transformed density p, we have $\mathbb{E}_{x\sim p}(X) = 0$ and $\mathbb{E}_{x\sim p}\left(XX^\top\right) = I$, i.e., zero mean and identity covariance; we call such a distribution *isotropic*. For any distribution with mean μ and covariance A, both well-defined, we can apply the transformation $A^{-\frac{1}{2}}(X-\mu)$ to get an isotropic distribution. Using this, we can reformulate the KLS conjecture as follows:

Conjecture (KLS, reformulated)**.** *For any logconcave density p in $\mathbb{R}^n$ with covariance matrix A, $\psi_p \gtrsim \|A\|_{\text{op}}^{-\frac{1}{2}}$. Equivalently, $\psi_p \gtrsim 1$ for any isotropic logconcave distribution p.*

1.2 Concentration of measure

One motivation to study isoperimetry is the phenomenon known as concentration of measure. This can be illustrated as follows: most of a Euclidean unit ball in $\mathbb{R}^n$ lies within distance $O(\frac{1}{n})$ of its boundary, and also within distance $O(\frac{1}{\sqrt{n}})$ of any central hyperplane. Most of a Gaussian lies in an annulus of thickness $O(1)$. For any subset of the sphere of measure $\frac{1}{2}$, the measure of points at distance at least $\sqrt{\frac{\log n}{n}}$ from the set is a vanishing fraction. These concentration phenomena are closely related to isoperimetry. For example, since the sphere has good isoperimetry (about $\sqrt{n}$), the boundary of any subset of measure $\frac{1}{2}$ is large, and summing up over all points within distance $\frac{1}{\sqrt{n}}$ gives a constant fraction of the entire sphere.



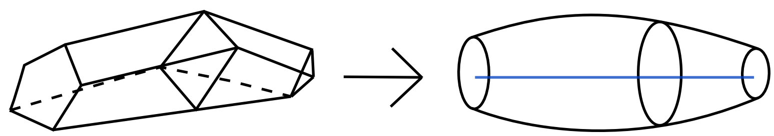
Concentration of measure

The relationship between isoperimetry and concentration runs deep with connections in both directions. In particular, the asymptotic behavior as the dimension grows is of interest in both cases and we will discuss this in more detail. We now review some basic definitions and properties of convex sets and logconcave distributions.

For two subsets $A, B \subseteq \mathbb{R}^n$ their Minkowski sum is $A+B = \{x+y : x \in A, y \in B\}$. The Brunn-Minkowski theorem says that if $A, B, A+B$ are measurable, then

$$\text{vol}(A+B)^{\frac{1}{n}} \geq \text{vol}(A)^{\frac{1}{n}} + \text{vol}(B)^{\frac{1}{n}}.$$

For the cross-sections of a convex body K orthogonal to a fixed vector u, it says that the volume function $v(t)$ along any direction is $\frac{1}{n-1}$-concave, i.e. $v(t)^{\frac{1}{n-1}}$ is concave. If we replace each cross-section with a ball of the same volume, the radius function is concave along u.



Brunn-Minkowski applied to convex bodies: the radius function is concave

A generalization of convex sets is logconcave functions (1.1). Their basic properties are summarized by the following classical lemma.

Lemma 2 (Dinghas; Prékopa; Leindler)**.** *The product, minimum and convolution of two logconcave functions is also logconcave; in particular, any linear transformation or marginal of a logconcave density is logconcave.*

Unlike convex sets, the set of logconcave distributions is closed under convolution. This is one of many reasons we work with this more general class of functions. The proof of this follows from an analog of the Brunn-Minkowski theorem for functions called the Prékopa-Leindler inequality: Let $\lambda \in [0,1]$, and three bounded functions $f, g, h : \mathbb{R}^n \to \mathbb{R}_+$ satisfy $f(\lambda x + (1-\lambda)y) \geq g(x)^\lambda h(y)^{1-\lambda}$ for any $x, y \in \mathbb{R}^n$. Then,

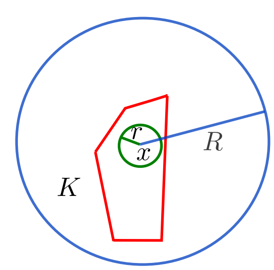
$$\int_{\mathbb{R}^n} f \geq \left(\int_{\mathbb{R}^n} g\right)^\lambda \left(\int_{\mathbb{R}^n} h\right)^{1-\lambda}.$$

1.3 The original motivation: A sampling algorithm

The study of algorithms has rich connections to high-dimensional convex geometry. It was the study of an algorithm for sampling that led to the KLS conjecture.

1.3.1 Model of computation

For algorithmic problems such as sampling, optimization and integration, the following general model of computation is standard. Convex bodies and logconcave functions are presented by *well-guaranteed* oracles. For a convex body $K \subset \mathbb{R}^n$, a well-guaranteed membership oracle is given by numbers $R \geq r > 0$, a point $x_0 \in K$ with the guarantee that $x_0 + rB_n \subseteq K \subseteq RB_n$ and an oracle that answers YES or NO to a query of the form "$x \in K$?" for any $x \in \mathbb{R}^n$. Another oracle of interest is a well-guaranteed *separation* oracle; it is given by the same parameters r, R (but no starting point in the set), and a stronger oracle: for a query "$x \in K$?", the oracle either answers YES or answers NO and provides a hyperplane that separates x from K. For an integrable function $f : \mathbb{R}^n \to \mathbb{R}_+$, a well-guaranteed function oracle is given by a bound on the norm of its center of gravity, upper and lower bounds on the eigenvalues of the covariance matrix with density proportional to f and an oracle that returns $f(x)$ for any $x \in \mathbb{R}^n$. The complexity of an algorithm is measured first in terms of the number of calls to the oracle and second in terms of the total number of arithmetic operations performed. An algorithm is considered efficient (and a problem *tractable*) if its complexity is bounded by a fixed polynomial in the dimension and other input parameters, which typically include an error parameter and a probability of failure parameter. We use $\widetilde{O}(\cdot)$ to suppress logarithmic factors in the leading expression and $O^*(\cdot)$ to suppress logarithmic factors as well as dependence on error parameters.



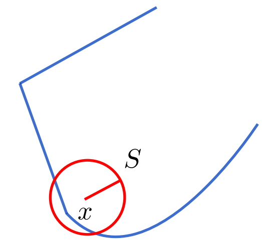
Well-guaranteed oracle

1.3.2 Sampling with a Markov chain

An important problem in the theory of algorithms is efficiently sampling high-dimensional sets and distributions. As we will presently see, sampling is closely related to an even more basic and ancient problem: estimating the volume (or integral).

Algorithms for sampling are based on Markov chains whose stationary distribution is the target distribution for the sampling problem. One such method is the *ball walk* [67] for sampling from the uniform distribution over a convex body (compact convex set), a particular discretization of Brownian motion. Start with some point in the body. At a current point x,

1. Pick a random point y in the ball of a fixed radius δ centered at x.
2. If y is in the body, go to y, otherwise stay at x.



Ball walk

This process converges to the uniform distribution over any compact, "well-connected" domain. But how efficient is it? To answer this question, we have to study its rate of convergence. This is a subject on its own with several general tools.

One way to bound the rate of convergence is via the smallest non-zero eigenvalue of the transition operator. If this eigenvalue is λ, then an appropriate notion of distance from the current distribution to the stationary distribution decreases by a multiplicative factor of $(1-\lambda)^t$ after t steps of a discrete-time Markov chain. Thus, the larger the spectral gap, the more rapid the convergence.

- But how to bound the spectral gap? There are many methods. In the context of high-dimensional continuous distributions, where the state space is not finite, one useful method is the notion of *conductance*. This can be viewed as the isoperimetry or expansion of the state space under the transition operator. More precisely, for a Markov chain defined by a state space Ω with transition operation $p(x \to y)$ and stationary distribution Q, the conductance of a measurable subset is

$$\phi(S) \stackrel{\text{def}}{=} \frac{\int_{y \notin S} \int_{x \in S} P(x \to y) Q(x) dx dy}{\min \{Q(S), Q(\Omega \setminus S)\}}$$

and the conductance of the entire Markov chain is

$$\phi \stackrel{\text{def}}{=} \inf_{S \subset \Omega} \phi(S).$$

The following general theorem, due to Jerrum and Sinclair [82] was extended to the continuous setting by Lovász and Simonivits [69].

Theorem 3 ([69])**.** *Let Q_t be the distribution of the current point after t steps of a Markov chain with stationary distribution Q and conductance at least ϕ, starting from initial distribution Q_0. Then, with $M = \sup_A \frac{Q_0(A)}{Q(A)}$,*

$$d_{TV}(Q_t, Q) \le \sqrt{M} \left(1 - \frac{\phi^2}{2}\right)^t$$

where $d_{TV}(Q_t, Q)$ is the total variation distance between Q_t and Q.

The mixing time of a Markov chain is related to its conductance via the following fundamental inequality.

$$\frac{1}{\phi} \lesssim \tau \lesssim \frac{1}{\phi^2} \log M \tag{1.3}$$

The conductance of the ball walk can be lower bounded by the product of two parameters. The first parameter is a mix of probability and geometry and asks for the minimum distance such that two points at this distance will have some constant overlap in their next-step distributions. The second is a purely geometry question, it is exactly the Cheeger constant of the stationary distribution. It turns out the first parameter can be estimated (for a precise statement, see Theorem 15), thereby reducing the problem of bounding the conductance of the Markov chain to bounding the Cheeger constant of the stationary distribution. This motivated the conjecture that the Cheeger constant for any isotropic logconcave density is in fact at least a universal constant, independent of the density and the dimension. If true, it would imply a bound of $O^*(n^2)$ on the mixing time of the ball walk from a warm start in an isotropic convex body, which is the best possible bound (it is tight for the isotropic hypercube).

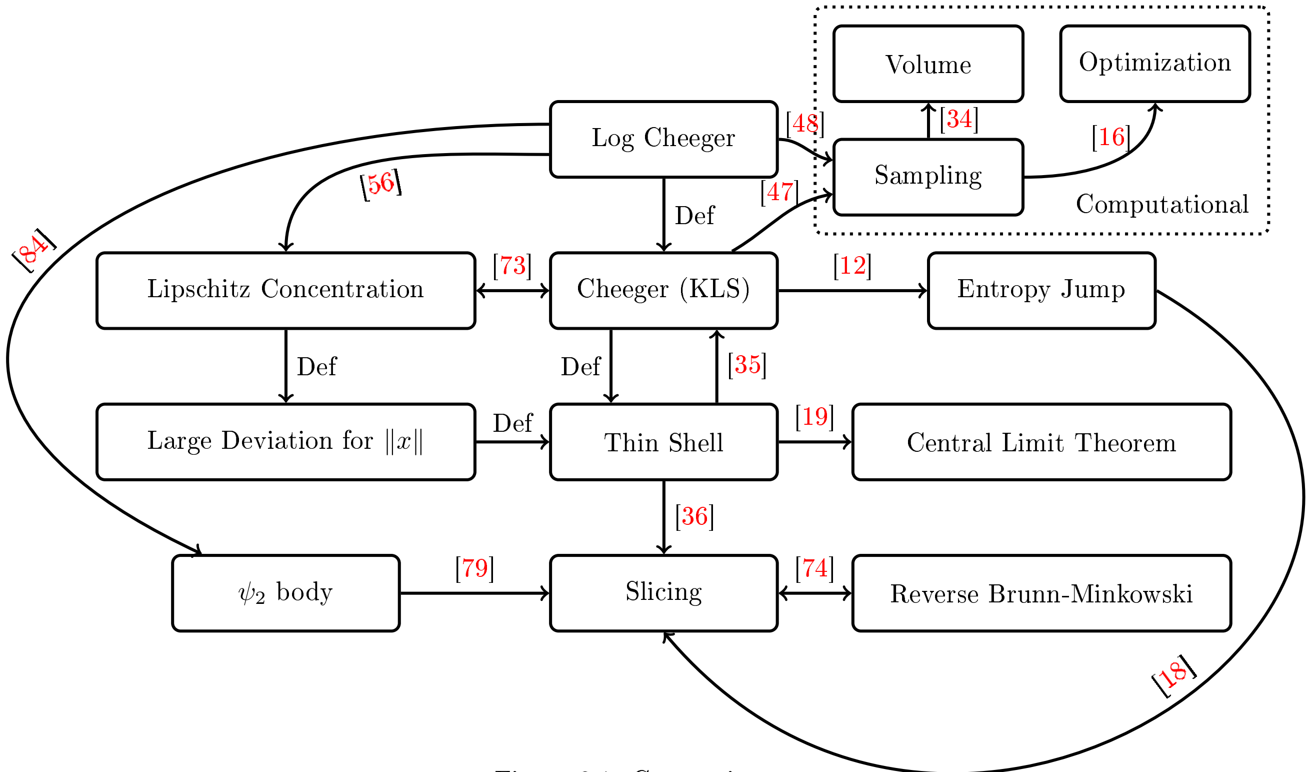


Figure 2.1: Connections

2 Connections

Here we discuss a wide array of implications of the conjecture, in geometry, probability and algorithms. For further details, we refer the reader to recent books on the topic [23, 5, 7].

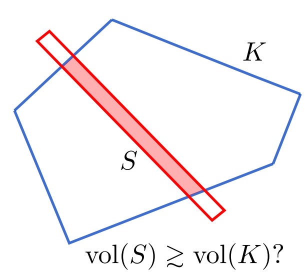
2.1 Geometry and Probability

The KLS conjecture implies the slicing conjecture and the thin-shell conjecture. Each of these has powerful and surprising consequences. We discuss them in order of the strength of the conjectures — slicing, thin-shell, KLS.

2.1.1 Slicing to anti-concentration

The slicing conjecture (a.k.a. the hyperplane conjecture) is one of the main open questions in convex geometry. It is implied by KLS conjecture. Ball first showed that a positive resolution of the KLS conjecture implies the slicing conjecture [10]. Eldan and Klartag [36] later gave a more refined quantitative relation (Theorem 10).

The conjecture says that any convex body in $\mathbb{R}^n$ of unit volume has a hyperplane section whose $(n-1)$-dimensional volume is at least a universal constant. Ball [11] gave the following equivalent conjecture for logconcave distributions.



Slicing Conjecture

Conjecture 4 (Slicing conjecture [22, 11])**.** *For any isotropic logconcave density p in $\mathbb{R}^n$, the isotropic (slicing) constant $L_p \stackrel{\text{def}}{=} p(0)^{\frac{1}{n}}$ is $O(1)$.*

Geometrically, this conjecture says that an isotropic logconcave distribution cannot have much mass around the origin. The best known bound is $L_p \lesssim n^{\frac{1}{4}}$ [50, 22].

Paouris showed that if the slicing conjecture is true, then a logconcave distribution satisfies a strong anti-concentration property (small ball probability).

Theorem 5 (Small ball probability [79])**.** *If slicing conjecture is true, for any isotropic logconcave density p in $\mathbb{R}^n$, we have*

$$\mathbb{P}_{x\sim p}\left(\|x\| \leq t\sqrt{n}\right) = O(t)^n$$

for all $0 \leq t \leq c$ for some universal constant c.

Paouris also showed that the inequality holds unconditionally with exponent $O(\sqrt{n})$ [79].

Another nice application of anti-concentration is to lower bound the entropy of a distribution.

Theorem 6 (Entropy of logconcave distribution [18])**.** *If the slicing conjecture is true, for any isotropic logconcave density p in $\mathbb{R}^n$, we have*

$$-O(n) \leq \mathbb{E}_{x\sim p} \log \frac{1}{p(x)} \leq O(n).$$

The Brunn-Minkowski inequality is not tight when applied to convex bodies that are large in different directions (e.g., $K = \{\varepsilon x^2 + \varepsilon^{-1} y^2 \leq 1\}$ and $T = \{\varepsilon^{-1} x^2 + \varepsilon y^2 \leq 1\}$ with tiny ε). The anti-concentration aspect of slicing also shows the following reverse inequality.

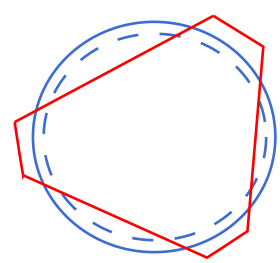
Theorem 7 (Reverse Brunn-Minkowski Inequalities [74])**.** *If slicing conjecture is true, then for any isotropic convex sets K and T, we have*

$$vol(K+T)^{1/n} \leq O(1)\left(vol(K)^{1/n} + vol(T)^{1/n}\right).$$

We remark that all these consequences of the slicing conjecture are in fact equivalent to the conjecture itself [32, 18, 21].

2.1.2 Thin shell to central limit theorem

The Central Limit Theorem says that a random marginal of a hypercube is approximately Gaussian. Brehm and Voigt asked if the same is true for convex sets. Anttila, Ball and Perissinaki [72] observed that this is true for any distribution on the sphere. Therefore, this also holds for any distributions with the norm $\|x\|$ concentrated at some value. Here, we state the version by Bobkov that holds for any distribution.



Thin Shell conjecture

Theorem 8 (Central Limit Theorem [19])**.** *Let μ be an isotropic probability on $\mathbb{R}^n$, which might not be logconcave. Assume that*

$$\mu\left(\left|\frac{\|x\|_2}{\sqrt{n}} - 1\right| \geq \varepsilon\right) \leq \varepsilon \tag{2.1}$$

for some $0 < \varepsilon < 1/3$. Let $g_\theta(s) = \mu(\{x^\top \theta = s\})$ and $g(s) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{s^2}{2})$. Then, for every $\delta > 0$, we have that

$$\mathbb{P}\left(\left\{\theta \in S^{n-1}: \max_{t\in\mathbb{R}}\left|\int_{-\infty}^{t} g_\theta(s)ds - \int_{-\infty}^{t} g(s)ds\right| \geq 2\delta + \frac{6}{\sqrt{n}} + 4\varepsilon\right\}\right) \leq c_1 \delta^{-\frac{3}{2}} \exp(-c_2\delta^4 n)$$

for some universal constants $c_1, c_2 > 0$.

So, it suffices to prove that $\|x\|$ is concentrated near $\sqrt{n}$ for any isotropic logconcave distribution. In a seminal work, Klartag proved that (2.1) holds with $\varepsilon \lesssim \log^{-\frac{1}{2}} n$ [51]. Shortly after, Fleury, Guédon and Paouris gave an alternative proof with $\varepsilon \lesssim \log^{-1/6} n \cdot (\log\log n)^2$ [40]. It is natural to ask for the optimal bound for (2.1). If the KLS conjecture is true, we can apply the conjecture for the sets $\{\|x\|_2 \leq r\}$ for different values of r and get the following conjecture, suggested in [72, 17].

Conjecture 9 (Thin shell conjecture). *For any isotropic logconcave density p in $\mathbb{R}^n$, the thin shell constant $\sigma_p \stackrel{\text{def}}{=} Var_{x\sim p}\left\Vert X\right\Vert_2$ is $O(1)$.*

The relation $\sigma_p \lesssim \psi_p^{-1}$ is an exercise. Eldan and Klartag showed that the slicing constant is bounded by the thin-shell constant (to within a universal constant).

Theorem 10 ([36]). *$L_p \lesssim \sup_p \sigma_p$ where the maximization is over all isotropic logconcave distribution.*

Since Klartag's bound on σ_p, there has been much effort to improve the bound (see Table 1). In a breakthrough, Eldan [35] showed that the thin shell conjecture is in fact equivalent to the KLS conjecture up a logarithmic factor (see Theorem 35).

Year/Authors	σ_p
2006/Klartag [51]	$\sqrt{n}/\sqrt{\log n}$
2006/Fleury-Guédon-Paouris [40]	$\sqrt{n}\frac{(\log\log n)^2}{\log^{1/6} n}$
2006/Klartag [52]	$n^{4/10}$
2010/Fleury [39]	$n^{3/8}$
2011/Guedon-Milman [44]	$n^{1/3}$
2016/Lee-Vempala [65]	$n^{1/4}$

Table 1: Progress on the thin shell bound.

2.1.3 Isoperimetry to concentration

After discussing two conjectures that are potentially weaker than KLS conjecture, we now move to some implications of the KLS conjecture itself.

The Poincaré constant of a measure is the minimum possible ratio of the expected squared gradient to the variance over smooth functions. Applying Cheeger's inequality [29] this constant is at least the square of the Cheeger constant. The reverse inequality holds for logconcave measures and was proved by Buser [26] (Ledoux [55] gave another proof).

Theorem 11 (Poincaré inequality [88, 29, 26, 55]). *For any isotropic logconcave density p in $\mathbb{R}^n$, we have that*

$$\zeta_p \stackrel{\text{def}}{=} \inf_{\text{smooth } g} \frac{\mathbb{E}_p\left(\left\Vert\nabla g(x)\right\Vert_2^2\right)}{Var_p\left(g(x)\right)} \approx \psi_p^2.$$

for any smooth g.

The Poincaré inequality is important in the study of partial differential equations. For example, the Poincaré constant governs exactly how fast the heat equation converges. For the logconcave setting, the choice of ℓ_2 norm is not important and it can be generalized as follows:

Theorem 12 (Generalized Poincaré inequality [73]). *For any isotropic logconcave density p in $\mathbb{R}^n$ and for all $1 \leq q < \infty$, we have that*

$$\mathbb{E}_{x\sim p}\left|\nabla g(x)\right| \gtrsim \frac{\psi_p}{q}\cdot\left(\mathbb{E}_{x\sim p}\left|g(x)-\mathbb{E}_{y\sim p}g(y)\right|^q\right)^{1/q}$$

for any smooth g.

Together with previous inequalities, we can summarize the relationships as follows: for any isotropic logconcave density p in $\mathbb{R}^n$,

$$L_p \lesssim \sup_p \sigma_p \quad \text{and} \quad \sigma_p \lesssim \frac{1}{\sqrt{\zeta_p}} \approx \frac{1}{\psi_p}$$

where the relation "$\lesssim$" hides only universal constants independent of the density p and the dimension n.

We next turn to concentration inequalities. The classical concentration theorem of Levy says that any Lipschitz function g on the sphere in $\mathbb{R}^n$ is concentrated near its mean (or median):

$$\mathbb{P}\left(|g(x) - \mathbb{E}_y g(y)| \geq t\right) \leq 2e^{-\frac{1}{2}t^2 n}.$$

The following theorem is an analogous statement for any logconcave density, and is due to Gromov and Milman.

Theorem 13 (Lipschitz concentration [42])**.** *For any L-Lipschitz function g in $\mathbb{R}^n$, and isotropic logconcave density p,*

$$\mathbb{P}_{x\sim p}\left(|g(x) - \mathbb{E}g| > L \cdot t\right) = e^{-\Omega(t\psi_p)}.$$

Milman [73] showed the reverse, namely that a Lipschtiz concentration inequality implies a lower bound on the Cheeger constant.

The next consequence is information-theoretic. Let X be a random variable from an n-dimensional distribution with a density p. Its entropy is $\mathrm{Ent}(X) = -\mathbb{E}(\log p) = -\int_{\mathbb{R}^n} p(x)\log p(x)\, dx$. The Shannon-Stam inequality says that for independent random vectors $X, Y \sim p$,

$$\mathrm{Ent}\left(\frac{X+Y}{\sqrt{2}}\right) \geq \mathrm{Ent}(X)$$

with equality only if p is Gaussian. Quantifying the increase in entropy has been a subject of investigation. Ball and Nguyen [12] proved the following bound on the entropy gap.

Theorem 14 (Entropy jump [12])**.** *Let p be an isotropic logconcave density and $X, Y \sim p$ in $\mathbb{R}^n$ and Z be drawn from a standard Gaussian in $\mathbb{R}^n$. Then,*

$$\mathrm{Ent}\left(\frac{X+Y}{\sqrt{2}}\right) - \mathrm{Ent}(X) \gtrsim \psi_p^2\left(\mathrm{Ent}(Z) - \mathrm{Ent}(X)\right).$$

2.2 Algorithms

In this section we discuss algorithmic connections of the KLS conjecture.

2.2.1 Sampling

The ball walk can be used to sample from any density using a Metropolis filter. To sample from the density Q, we repeat the following: at a point x,

1. Pick a random point y in the δ-ball centered at x.
2. Go to y with probability $\min\left\{1, \frac{Q(y)}{Q(x)}\right\}$.

If the resulting Markov chain is ergodic, the current distribution approaches a unique stationary distribution. The complexity of sampling depends on the *rate* of convergence to stationarity. For an isotropic logconcave distribution, the rate of convergence of the ball walk from a *warm* starting distribution is bounded in terms of the Cheeger constant. A starting distribution Q_0 is *warm* with respect to the stationary distribution Q if $\mathbb{E}_{x\sim Q}\frac{Q_0(x)}{Q(x)}$ is bounded by a constant.

Theorem 15 ([47])**.** *For an isotropic logconcave density p, the ball walk mixes from a warm start in $O^*\left(n^2/\psi_p^2\right)$ steps.*

2.2.2 Sampling to Convex Optimization

Sampling can be used to efficiently implement a basic algorithm for convex optimization given a separation oracle — the *cutting plane* method. Convex optimization can be reduced to convex feasibility by including the objective function as a constraint and doing a binary search on its value. To solve the feasibility problem, the method maintains a convex set containing K, starting with the ball of radius R which is guaranteed to contain K. At each step it queries the centroid of the set. If infeasible, it uses the violated inequality given by the separation oracle to restrict the set. The basis of this method is the following theorem of Grunbaum.

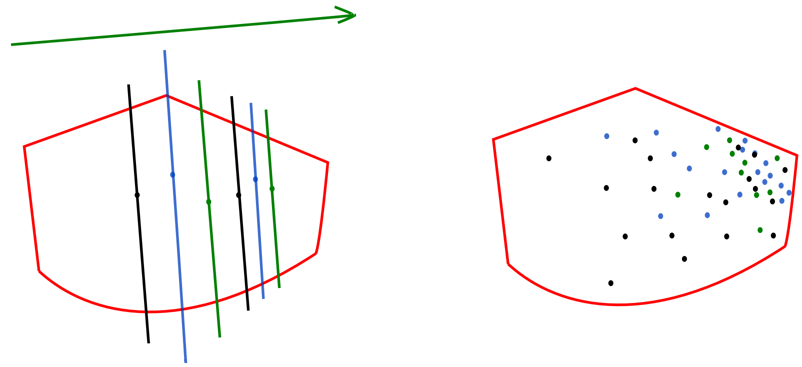


Figure 2.2: Centroid cuts vs Simulated annealing for optimization

Theorem 16 ([43]). *For any convex body K, for any halfspace H containing the centroid of K,* $\mathrm{vol}(H \cap K) \geq \frac{1}{e}\mathrm{vol}(K)$.

Thus, the volume of the set maintained decreases by a constant factor in each iteration and the number of iterations is $O(n \log \frac{R}{r})$, which is asymptotically the best possible. However, there is one important difficulty, namely computing the center of gravity, even of an explicit polytope [33], is a computationally intractable problem (#P-hard). The next theorem uses sampling to get around this, while keeping the same asymptotic complexity.

Theorem 17 ([16]). *Let $X = \frac{1}{m}\sum_{i=1}^{m} X_i$ where X_i are drawn i.i.d. from a logconcave density p in $\mathbb{R}^n$. Then, for any halfspace H containing X,*

$$\mathbb{E}\left(\int_H p\right) \geq \frac{1}{e} - \sqrt{\frac{n}{m}}.$$

Thus, for a convex body, using the average of $m = O(n)$ samples is an effective substitute for the center of gravity.

While this method achieves the best possible oracle complexity and (an impractically high) polynomial number of arithmetic operations, the work of Lee, Sidford and Wong [60] shows how to reduce the overall arithmetic complexity to $\widetilde{O}(n^3)$. Their general method also leads to the current fastest algorithms for semi-definite programming and submodular function minimization.

Convex optimization given only a membership oracle can also be reduced to sampling, via the method known as simulated annealing. It starts with a uniform distribution over the feasible set, then gradually focuses the distribution on near-optimal points. A canonical way to minimize the linear function $c^\top x$ over a convex body K is to use a sequence of Boltzmann-Gibbs distributions with density proportional to $e^{-\alpha c^\top x}$ for points in K, with α starting close to zero and gradually increasing it. A random point drawn from this density satisfies

$$\mathbb{E}\left(c^\top x\right) \leq \min_K c^\top x + \frac{n}{\alpha}.$$

Thus, sampling from the density with $\alpha = n/\epsilon$ gives an additive ϵ error approximation. In [45] it is shown how to make this method efficient, using a sequence of only $\widetilde{O}(\sqrt{n})$ distributions.

The method can be viewed as a special case of a more general family of algorithms referred to as the interior-point method [76]. The latter method works by minimizing the sum of the desired objective function with a diminishing multiple of a smooth convex function which blows up at the boundary. Thus the optimum points to this modified objective for any positive value of the multiplier are in the interior of the convex body. The path followed by the method as a function of the multiplier is called the central path. There are several interesting choices of the smooth convex functions, including the logarithmic barrier, volumetric barrier, universal barrier and entropic barrier. The last two of these achieve the optimal rate in terms of the dimension for arbitrary convex bodies. This rate, i.e., number of steps to reduce the distance to optimality

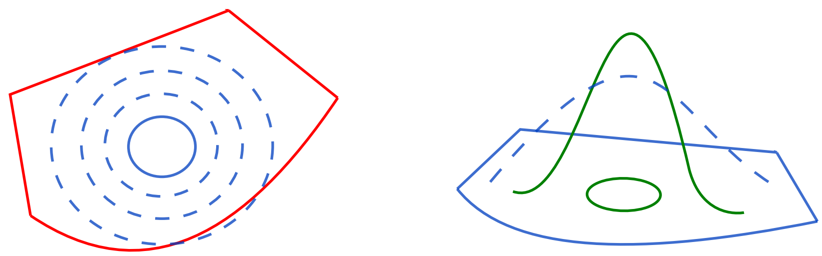


Figure 2.3: DFK vs Simulated Annealing

by a constant factor, is $O(\sqrt{n})$, the same as for simulated annealing. This is not a coincidence — the path of the centroid as a function of α corresponds exactly to the central path taken by the interior-point method using the entropic barrier [24, 1]. The universal and entropic barriers are not the most efficient for the important case of linear programming — each step can be implemented in polynomial time, but is n^4 or higher. For a linear program given by m inequalities, the logarithmic barrier uses $\widetilde{O}(\sqrt{m})$ phases, with each phase requiring the solution of a linear system. Lee and Sidford have proposed a barrier that takes only $\widetilde{O}(\sqrt{n})$ iterations and gives the currest fastest method for solving linear programs [58, 59].

Optimization based on sampling has various robustness properties (e.g., it can be applied to stochastic optimization [15, 38] and regret minimization [75, 25]), and continues to be an active research topic.

2.2.3 Sampling to Volume computation and Integration

Sampling is the core of efficient volume computation and integration. The main idea for the latter problems is to sample a sequence of logconcave distributions, starting with one that is easy to integrate and ending with the function whose integral is desired. This process, known as *simulated annealing* can be expressed as the following telescoping product:

$$\int_{\mathbb{R}^n} f = \int f_0 \frac{\int f_i}{\int f_0} \frac{\int f_2}{\int f_3} \cdots \frac{\int f_m}{\int f_{m-1}}$$

where $f_m = f$. Each ratio $\int f_{i+1} / \int f_i$ is the expectation of the estimator $Y = \frac{f_{i+1}(X)}{f_i(X)}$ for X drawn from the density proportional to f_i. What is the optimal sequence of interpolating functions to use? The celebrated polynomial-time algorithm of Dyer, Frieze and Kannan [34] used the uniform distribution on a sequence of convex bodies, starting with the ball contained inside the input body K. Each body in the sequence is a ball intersected with the given convex body K: $K_i = 2^{\frac{i}{n}} rB \cap K$. The length of the sequence is $m = O(n \log \frac{R}{r})$ so that the final body is just K. A variance bound shows that $O(m/\epsilon^2)$ samples per distribution suffice to get an overall $1+\epsilon$ multiplicative error approximation with high probability. The total number of samples is $O^*(m^2) = O^*(n^2)$ and the complexity of the resulting algorithm is $O^*(n^5)$ as shown in [47]. Table 2 below summarizes progress on the volume problem over the past three decades. Besides improving the complexity of volume computation, each step has typically resulted in new techniques. For more details, we refer the reader to surveys on the topic [81, 87].

Lovász and Vempala [70] improved on [47] by sampling from a sequence of nonuniform distributions. The densities in the sequence have the form $f_i(x) \propto \exp(-\alpha_i \|x\|)\chi_K(x)$ or $f_i(x) \propto \exp(-\alpha_i \|x\|^2)\chi_K(x)$. Then the ratio of two consecutive integrals is the expectation of the following estimator:

$$Y = \frac{f_{i+1}(X)}{f_i(X)}.$$

Year/Authors	New ingredients	Steps
1989/Dyer-Frieze-Kannan [34]	Everything	n^{23}
1990/Lovász-Simonovits [68]	Better isoperimetry	n^{16}
1990/Lovász [67]	Ball walk	n^{10}
1991/Applegate-Kannan [6]	Logconcave sampling	n^{10}
1990/Dyer-Frieze [33]	Better error analysis	n^{8}
1993/Lovász-Simonovits [69]	Localization lemma	n^{7}
1997/Kannan-Lovász-Simonovits [47]	Speedy walk, isotropy	n^{5}
2003/Lovász-Vempala [70]	Annealing, hit-and-run	n^{4}
2015/Cousins-Vempala [31] (well-rounded)	Gaussian Cooling	n^{3}
2017/Lee-Vempala (polytopes)	Hamiltonian Walk	$mn^{\frac{2}{3}}$

Table 2: The complexity of volume estimation, each step uses $\widetilde{O}(n)$ bit of randomness.

They showed that the coefficient α_i (inverse "temperature") can be changed by a factor of $(1+\frac{1}{\sqrt{n}})$, which implies that $m=\widetilde{O}(\sqrt{n})$ phases suffice, and the total number of samples is only $O^*(n)$. This is perhaps surprising since the ratio of the initial integral to the final is typically $n^{\Omega(n)}$. Even though the algorithm uses only $\widetilde{O}(\sqrt{n})$ phases, and hence estimates a ratio of $n^{\tilde{\Omega}(\sqrt{n})}$ in one or more phases, the variance of the estimator is bounded in every phase.

Theorem 18 ([70])**.** *The volume of a convex body in $\mathbb{R}^n$ (given by a membership oracle) can be computed to relative error ε using $\widetilde{O}(n^4/\varepsilon^2)$ oracle queries and $\widetilde{O}(n^2)$ arithmetic operations per query.*

The LV algorithm has two parts. In the first it finds a transformation that puts the body in near-isotropic position. The complexity of this part is $\widetilde{O}(n^4)$. In the second part, it runs the annealing schedule, while maintaining that the distribution being sampled is *well-rounded*, a weaker condition than isotropy. Well-roundedness requires that a level set of measure $\frac{1}{8}$ contains a constant-radius ball and the trace of the covariance (expected squared distance of a random point from the mean) to be bounded by $O(n)$, so that R/r is effectively $O(\sqrt{n})$. To achieve the complexity guarantee for the second phase, it suffices to use the KLS bound of $\psi_p \gtrsim n^{-\frac{1}{2}}$. Connecting improvements in the Cheeger constant directly to the complexity of volume computation is an open question. To apply improvements in the Cheeger constant, one would need to replace well-roundedness with (near-)isotropy and maintain that. However, maintaining isotropy appears to be much harder — possibly requiring a sequence of $\Omega(n)$ distributions and $\Omega(n)$ samples from each, providing no gain over the current complexity of $O^*(n^4)$ even if the KLS conjecture turns out to be true.

Cousins and Vempala [31] gave a faster algorithm for well-rounded convex bodies (any isotropic logconcave density satisfies $\frac{R}{r}=O(\sqrt{n})$ and is well-rounded). Their algorithm, called Gaussian cooling, is significantly simpler, crucially utilizes the fact that the KLS conjecture holds for a Gaussian density restricted by any convex body (Theorem 25), and completely avoids computing an isotropic transformation.

Theorem 19 ([31])**.** *The volume of a well-rounded convex body, i.e., with $R/r=O^*(\sqrt{n})$, can be computed using $O^*(n^3)$ oracle calls.*

We note that the covariance matrix of any logconcave density can be computed efficiently from only a linear in the dimension number of samples. This question of sample complexity was also motivated by the study of volume computation.

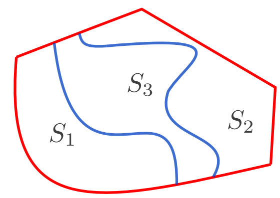
Theorem 20 ([2, 3, 83])**.** *Let $Y=\frac{1}{m}\sum_{i=1}^{m}X_iX_i^\top$ where $X_1,\ldots,X_m$ are drawn from an isotropic logconcave density p. If $m\gtrsim\frac{n}{\varepsilon^2}$, then $\|Y-I\|_{\mathrm{op}}\le\varepsilon$ with high probability.*

3 Proof techniques

Classical proofs of isoperimetry for special distributions are based on different types of symmetrization that effectively identify the extremal subsets. Bounding the Cheeger constant for general convex bodies and logconcave densities is more complicated since the extremal sets can be nonlinear and hard to describe precisely, due to the trade-off between minimizing the boundary measure of a subset and utilizing as much of the "external" boundary as possible. The main technique to prove bounds in the general setting has been *localization*, a method to reduce inequalities in high dimension to inequalities in one dimension. We now describe this technique with a few applications.

3.1 Localization

We will sketch a proof of the following theorem to illustrate the use of localization. This theorem was also proved by Karzanov and Khachiyan [49] using a different, more direct approach.



Euclidean isoperimetry

Theorem 21 ([33, 68, 49]). *Let f be a logconcave function whose support has diameter D and let π_f be the induced measure. Then for any partition of $\mathbb{R}^n$ into measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{2d(S_1,S_2)}{D}\min\{\pi_f(S_1), \pi_f(S_2)\}.$$

Before discussing the proof, we note that there is a variant of this result in the Riemannian setting.

Theorem 22 ([66]). *If $K \subset (M, g)$ is a locally convex bounded domain with smooth boundary, diameter D and $Ric_g \geq 0$, then the Poincaré constant is at least $\frac{\pi^2}{4D^2}$, i.e., for any g with $\int g = 0$, we have that*

$$\int |\nabla g(x)|^2 \, dx \geq \frac{\pi^2}{4D^2} \int g(x)^2 dx.$$

For the case of convex bodies in $\mathbb{R}^n$, this result is equivalent to Theorem 21 up to a constant. One benefit of localization is that it does not require a carefully crafted potential. Localization has recently been generalized to Riemannian setting [53]. The origins of this method were in a paper by Payne and Weinberger [80].

We begin the proof of Theorem 21. For a proof by contradiction, let us assume the converse of its conclusion, i.e., for some partition S_1, S_2, S_3 of $\mathbb{R}^n$ and logconcave density f, assume that

$$\int_{S_3} f(x)\,dx < C\int_{S_1} f(x)\,dx \quad \text{and} \quad \int_{S_3} f(x)\,dx < C\int_{S_2} f(x)\,dx$$

where $C = 2d(S_1, S_2)/D$. This can be reformulated as

$$\int_{\mathbb{R}^n} g(x)\,dx > 0 \quad \text{and} \quad \int_{\mathbb{R}^n} h(x)\,dx > 0 \tag{3.1}$$

where

$$g(x) = \begin{cases} Cf(x) & \text{if } x \in S_1, \\ 0 & \text{if } x \in S_2, \\ -f(x) & \text{if } x \in S_3. \end{cases} \quad \text{and} \quad h(x) = \begin{cases} 0 & \text{if } x \in S_1, \\ Cf(x) & \text{if } x \in S_2, \\ -f(x) & \text{if } x \in S_3. \end{cases}$$

These inequalities are for functions in $\mathbb{R}^n$. The next lemma will help us analyze them.

Lemma 23 (Localization Lemma [46]). *Let $g, h : \mathbb{R}^n \to \mathbb{R}$ be lower semi-continuous integrable functions such that*

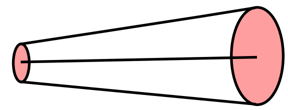
$$\int_{\mathbb{R}^n} g(x)\,dx > 0 \quad \textit{and} \quad \int_{\mathbb{R}^n} h(x)\,dx > 0.$$

Then there exist two points $a, b \in \mathbb{R}^n$ and an affine function $\ell : [0,1] \to \mathbb{R}_+$ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a+tb)\,dt > 0 \quad \textit{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a+tb)\,dt > 0.$$

The points a, b represent an interval and one may think of $\ell(t)^{n-1}$ as proportional to the cross-sectional area of an infinitesimal cone.

The lemma says that over this cone truncated at a and b, the integrals of g and h are positive. Also, without loss of generality, we can assume that a, b are in the union of the supports of g and h.



Truncated cone

Proof outline. The main idea is the following. Let H be any halfspace such that

$$\int_H g(x)\,dx = \frac{1}{2}\int_{\mathbb{R}^n} g(x)\,dx.$$

Let us call this a bisecting halfspace. Now either

$$\int_H h(x)\,dx > 0 \quad \text{or} \quad \int_{\mathbb{R}^n\setminus H} h(x)\,dx > 0.$$

Thus, either H or its complementary halfspace will have positive integrals for both g and h, reducing the domain of the integrals from $\mathbb{R}^n$ to a halfspace. If we could repeat this, we might hope to reduce the dimensionality of the domain. For any $(n-2)$-dimensional affine subspace L, there is a bisecting halfspace containing L in its bounding hyperplane. To see this, let H be a halfspace containing L in its boundary. Rotating H about L we get a family of halfspaces with the same property. This family includes H', the complementary halfspace of H. The function $\int_H g - \int_{\mathbb{R}^n\setminus H} g$ switches sign from H to H'. Since this is a continuous family, there must be a halfspace for which the function is zero.

If we take all $(n-2)$-dimensional affine subspaces defined by $\{x \in \mathbb{R}^n : x_i = r_1, x_j = r_2\}$ where r_1, r_2 are rational, then the intersection of all the corresponding bisecting halfspaces is a line or a point (by choosing only rational values for x_i, we are considering a countable intersection). To see why it is a line or a point, assume we are left with a two or higher dimensional set. Since the intersection is convex, there is a point in its interior with at least two coordinates that are rational, say $x_1 = r_1$ and $x_2 = r_2$. But then there is a bisecting halfspace H that contains the affine subspace given by $x_1 = r_1$, $x_2 = r_2$ in its boundary, and so it properly partitions the current set.

Thus the limit of this bisection process is a function supported on an interval (which could be a single point), and since the function itself is a limit of convex sets (intersections of halfspaces) containing this interval, it is a limit of a sequence of concave functions and is itself concave, with positive integrals. Simplifying further from concave to linear takes quite a bit of work. For the full proof, we refer the reader to [69]. $\square$

Going back to the proof sketch of Theorem 21, we can apply the localization lemma to get an interval $[a, b]$ and an affine function ℓ such that

$$\int_0^1 \ell(t)^{n-1} g((1-t)a + tb)\,dt > 0 \quad \text{and} \quad \int_0^1 \ell(t)^{n-1} h((1-t)a + tb)\,dt > 0. \tag{3.2}$$

The functions g, h as we have defined them are not lower semi-continuous. However, this can be addressed by expanding S_1 and S_2 slightly so as to make them open sets, and making the support of f an open set. Since we are proving strict inequalities, these modifications do not affect the conclusion.

Let us partition $[0, 1]$ into Z_1, Z_2, Z_3 as follows:

$$Z_i = \{t \in [0,1] : (1-t)a + tb \in S_i\}.$$

Note that for any pair of points $u \in Z_1, v \in Z_2$, $|u - v| \geq d(S_1, S_2)/D$. We can rewrite (3.2) as

$$\int_{Z_3} \ell(t)^{n-1} f((1-t)a + tb)\,dt < C \int_{Z_1} \ell(t)^{n-1} f((1-t)a + tb)\,dt$$

and

$$\int_{Z_3} \ell(t)^{n-1} f((1-t)a + tb)\,dt < C \int_{Z_2} \ell(t)^{n-1} f((1-t)a + tb)\,dt.$$

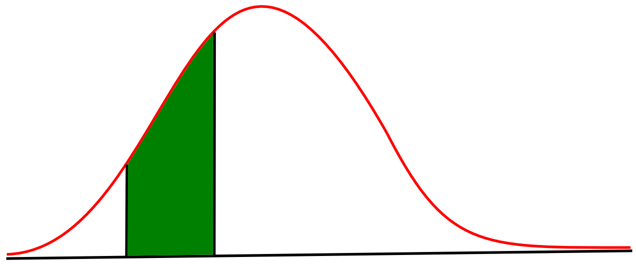


Figure 3.1: One-dimensional isoperimetry

The functions f and $\ell(\cdot)^{n-1}$ are both logconcave, so $F(t) = \ell(t)^{n-1} f((1-t)a + tb)$ is also logconcave. We get,

$$\int_{Z_3} F(t)\, dt < C \min \left\{ \int_{Z_1} F(t)\, dt, \int_{Z_2} F(t)\, dt \right\}. \tag{3.3}$$

Now consider what Theorem 21 asserts for the function $F(t)$ over the interval $[0,1]$ and the partition Z_1, Z_2, Z_3:

$$\int_{Z_3} F(t)\, dt \geq 2d(Z_1, Z_2) \min \left\{ \int_{Z_1} F(t)\, dt, \int_{Z_2} F(t)\, dt \right\}. \tag{3.4}$$

We have substituted 1 for the diameter of the interval $[0,1]$. Also, $2d(Z_1, Z_2) \geq 2d(S_1, S_2)/D = C$. Thus, Theorem 21 applied to the function $F(t)$ contradicts (3.3) and to prove the theorem in general, and it suffices to prove it in the one-dimensional case. A combinatorial argument reduces this to the case when each Z_i is a single interval. Proving the resulting inequality up to a factor of 2 is a simple exercise and uses only the unimodality of F. The improvement to the tight bound requires one-dimensional logconcavity. This completes the proof of Theorem 21.

The localization lemma has been used to prove a variety of isoperimetric inequalities. The next theorem is a refinement of Theorem 21, replacing the diameter by the square-root of the expected squared distance of a random point from the mean. For an isotropic distribution this is an improvement from n to $\sqrt{n}$. This theorem was proved by Kannan-Lovász-Simonovits in the same paper in which they proposed the KLS conjecture.

Theorem 24 ([46])**.** *For any logconcave density p in $\mathbb{R}^n$ with covariance matrix A, the KLS constant satisfies*

$$\psi_p \gtrsim \frac{1}{\sqrt{\operatorname{Tr}(A)}}.$$

The next theorem shows that the KLS conjecture is true for an important family of distributions. The proof is again by localization [30], and the one-dimensional inequality obtained is a Brascamp-Lieb Theorem. We note that the same theorem can be obtained by other means [57, 35].

Theorem 25. *Let $h(x) = f(x)e^{-\frac{1}{2}x^\top Bx} / \int f(y)e^{-\frac{1}{2}y^\top By} dy$ where $f : \mathbb{R}^n \to \mathbb{R}_+$ is an integrable logconcave function and B is positive definite. Then h is logconcave and for any measurable subset S of $\mathbb{R}^n$,*

$$\frac{h(\partial S)}{\min\left\{h(S), h\left(\mathbb{R}^n \setminus S\right)\right\}} \gtrsim \frac{1}{\|B^{-1}\|_{\mathrm{op}}^{\frac{1}{2}}}.$$

In other words, the expansion of h is $\Omega(\left\|B^{-1}\right\|_{\mathrm{op}}^{-\frac{1}{2}})$.

The analysis of the Gaussian Cooling algorithm for volume computation [31] uses localization. Kannan, Lovász and Montenegro [48] used localization to prove the following bound on the log-Cheeger constant, a quantity that is asymptotically the square-root of the log-Sobolev constant [55], and which we will discuss in the next section.

Theorem 26 ([48]). *The log-Cheeger constant of any isotropic logconcave density with support of diameter D satisfies $\kappa_p \gtrsim \frac{1}{D}$ where*

$$\kappa_p = \inf_{S \subseteq \mathbb{R}^n} \frac{p(\partial S)}{\min\left\{p(S)\log\frac{1}{p(S)}, p(\mathbb{R}^n \setminus S)\log\frac{1}{p(\mathbb{R}^n \setminus S)}\right\}}.$$

Next we mention an application to the anti-concentration of polynomials. This is a corollary of a more general result by Carbery and Wright.

Theorem 27 ([27]). *Let q be a degree d polynomial in $\mathbb{R}^n$. Then for a convex body $K \subset \mathbb{R}^n$ of volume 1, any $\epsilon > 0$, and x drawn uniform from K,*

$$\Pr_{x \sim K}\left(|q(x)| \le \epsilon \max_K |q(x)|\right) \lesssim \epsilon^{\frac{1}{d}} n$$

We conclude this section with a nice interpretation of the localization lemma by Fradelizi and Guedon. They also give a version that extends localization to multiple inequalities.

Theorem 28 (Reformulated Localization Lemma [41]). *Let K be a compact convex set in $\mathbb{R}^n$ and f be an upper semi-continuous function. Let P_f be the set of logconcave distributions μ supported by K satisfying $\int f d\mu \ge 0$. The set of extreme points of $convP_f$ is exactly:*

1. *the Dirac measure at points x such that $f(x) \ge 0$, or*
2. *the distributions v satisfies*
   (a) *density function is of the form e^{ℓ} with linear ℓ,*
   (b) *support equals to a segment $[a, b] \subset K$*
   (c) $\int f dv = 0$
   (d) $\int_a^x f dv > 0$ *for* $x \in (a, b)$ *or* $\int_x^b f dv > 0$ *for* $x \in (a, b)$.

Since we know the maximizer of any convex function is at extreme points, this shows that one can optimize $\max_{\mu \in P_f} \Phi(\mu)$ for any convex Φ by checking Dirac measures and log-affine functions!

3.2 Stochastic localization

We now describe a variant of localization that is the key idea behind recent progress on the KLS conjecture. Consider a subset E with measure 0.5 according to a logconcave density (it suffices to consider such subsets to bound the isoperimetric constant [73]). In standard localization, we repeatedly bisect space using a hyperplane that preserves the relative measure of E. The limit of this process is a partition into 1-dimensional logconcave measures ("needles"), for which inequalities are easier to prove. This approach runs into difficulties for proving the KLS conjecture. While the original measure might be isotropic, the one-dimensional measures could, in principle, have variances roughly equal to the trace of the original covariance (i.e., long thin needles), for which the Cheeger constant is much smaller. However, if we pick the bisecting hyperplanes randomly, it seems unlikely that we get such long thin needles. In a different line of work, Klartag introduced a "tilt" operator, i.e., multiplying a density by an exponential of the form $f(x) \propto e^{-c^\top x}$ along a fixed vector c, and used it in his paper improving the slicing constant [50, 54]. A stochastic version of localization combining both these aspects, i.e., apprpaching needles via tilt operators, was discovered by Eldan [35].

Stochastic localization can be viewed as the continuous-time process, where at each step, we pick a random direction and multiply the current density with a linear function along the chosen direction. Thus, the discrete bisection step is replaced by infinitesimal steps, each of which is a renormalization with a linear function in a random direction. One might view this informally as an averaging over infinitesimal needles. The discrete time equivalent would be $p_{t+1}(x) = p_t(x)(1 + \sqrt{h}(x - \mu_t)^\top w)$ for a sufficiently small h and random Gaussian vector w. Using the approximation $1 + y \sim e^{y - \frac{1}{2}y^2}$, we see that over time this process introduces a negative quadratic factor in the exponent, which will be the Gaussian factor. As time tends to ∞, the distribution tends to a more and more concentrated Gaussian and eventually a delta function, at

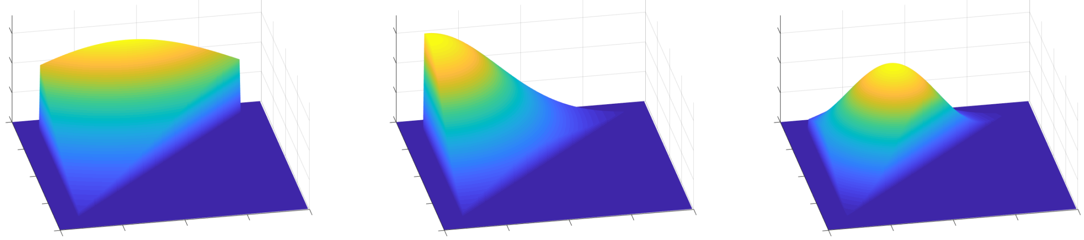


Figure 3.2: Stochastic localization

which point any subset has measure either 0 or 1. The idea of the proof is to stop at a time that is large enough to have a strong Gaussian factor in the density, but small enough to ensure that the measure of a set is not changed by more than a constant. Over time, the density can be viewed as a spherical Gaussian times a logconcave function, with the Gaussian gradually reducing in variance. When the Gaussian becomes sufficiently small in variance, then the overall distribution has good isoperimetric coefficient, determined by the inverse of the Gaussian standard deviation (Theorem 25). An important property of the infinitesimal change at each step is *balance* – the density at time t is a martingale and therefore the expected measure of any subset is the same as the original measure. Over time, the measure of a set E is a random quantity that deviates from its original value of $\frac{1}{2}$ over time. The main question then is: what direction to use at each step so that (a) the measure of E remains bounded and (b) the Gaussian factor in the density eventually has small variance.

3.2.1 A stochastic process and its properties

Given a distribution with logconcave density $p(x)$, we start at time $t=0$ with this distribution and at each time $t>0$, we apply an infinitesimal change to the density. To make some proofs easier, one may assume that the support of p is contained in a ball of radius n because there is only exponentially small probability outside this ball, at most $e^{-\Omega(n)}$. Let dW_t be the infinitesimal Wiener process.

Definition 29. Given a logconcave distribution p, we define the following stochastic differential equation:

$$c_0 = 0, \quad dc_t = dW_t + \mu_t dt, \tag{3.5}$$

where the probability distribution p_t and its mean μ_t are defined by

$$p_t(x) = \frac{e^{c_t^\top x - \frac{t}{2}\|x\|_2^2} p(x)}{\int_{\mathbb{R}^n} e^{c_t^\top y - \frac{t}{2}\|y\|_2^2} p(y) dy}, \quad \mu_t = \mathbb{E}_{x\sim p_t} x.$$

We will presently explain why p_t takes this form with a Gaussian component. Before we do that, we note that the process can be generalized using a "control" matrix C_t at time t. This is a positive definite matrix that could, for example, be used to adapt the process to the covariance of the current distribution. At time t, the covariance matrix is

$$A_t \stackrel{\text{def}}{=} \mathbb{E}_{x\sim p_t}(x-\mu_t)(x-\mu_t)^\top$$

The control matrix is incorporated in the following more general version of (3.5):

$$c_0 = 0, \quad dc_t = C_t^{1/2} dW_t + C_t \mu_t dt,$$
$$B_0 = 0, \quad dB_t = C_t dt,$$

where the probability distribution p_t is now defined by

$$p_t(x) = \frac{e^{c_t^\top x - \frac{1}{2}\|x\|_{B_t}^2} p(x)}{\int_{\mathbb{R}^n} e^{c_t^\top y - \frac{1}{2}\|y\|_{B_t}^2} p(y) dy}.$$

When C_t is a Lipschitz function with respect to c_t, μ_t, A_t and t, standard theorems (e.g., [77, Sec 5.2]) show the existence and uniqueness of the solution in time $[0,T]$ for any $T>0$.

We will now focus on the case $C_t = I$ and hence $B_t = tI$. The lemma below says that the stochastic process is the same as continuously multiplying $p_t(x)$ by a random infinitesimally small linear function.

Lemma 30. *We have that $dp_t(x) = (x-\mu_t)^\top dW_t \cdot p_t(x)$ for any $x \in \mathbb{R}^n$.*

3.2.2 Alternative definition of the process

Here, we use $dp_t(x) = (x-\mu_t)^\top dW_t p_t(x)$ as the definition of the process and show how the Gaussian term $-\frac{t}{2}\|x\|_2^2$ emerges. To compute $d\log p_t(x)$, we first explain how to apply the chain rule (Itô's formula) for a stochastic processes. Given real-valued stochastic processes x_t and y_t, the quadratic variations $[x]_t$ and $[x,y]_t$ are real-valued stochastic processes defined by

$$[x]_t = \lim_{|P|\to 0} \sum_{n=1}^{\infty} \left(x_{\tau_n} - x_{\tau_{n-1}}\right)^2,$$
$$[x,y]_t = \lim_{|P|\to 0} \sum_{n=1}^{\infty} \left(x_{\tau_n} - x_{\tau_{n-1}}\right)\left(y_{\tau_n} - y_{\tau_{n-1}}\right),$$

where $P = \{0 = \tau_0 \le \tau_1 \le \tau_2 \le \cdots \uparrow t\}$ is a stochastic partition of the non-negative real numbers, $|P| = \max_n(\tau_n - \tau_{n-1})$ is called the *mesh* of P and the limit is defined using convergence in probability. Note that $[x]_t$ is non-decreasing with t and $[x,y]_t$ can be defined as

$$[x,y]_t = \frac{1}{4}\left([x+y]_t - [x-y]_t\right).$$

For example, if the processes x_t and y_t satisfy the SDEs

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t \quad \text{and} \quad y_t = \nu(y_t)dt + \eta(y_t)dW_t$$

where W_t is a Wiener process, we have that

$$[x]_t = \int_0^T \sigma^2(x_s)ds, \quad [x,y]_t = \int_0^T \sigma(x_s)\eta(y_s)ds \quad \text{and} \quad d[x,y]_t = \sigma(x_t)\eta(y_t)dt.$$

For vector-valued SDEs

$$dx_t = \mu(x_t)dt + \Sigma(x_t)dW_t \quad \text{and} \quad dy_t = \nu(y_t)dt + M(y_t)dW_t$$

we have that

$$[x^i, x^j]_t = \int_0^T (\Sigma(x_s)\Sigma^\top(x_s))_{ij}ds \quad \text{and} \quad d[x^i, y^j]_t = \int_0^T (\Sigma(x_s)M^\top(y_s))_{ij}ds.$$

Lemma 31 (Itô's formula)**.** *Let x be a semimartingale and f be a twice continuously differentiable function, then*

$$df(x_t) = \sum_i \frac{df(x_t)}{dx^i}dx^i + \frac{1}{2}\sum_{i,j}\frac{d^2 f(x_t)}{dx^i dx^j}d[x^i, x^j]_t.$$

We can now compute the derivative $d\log p_t(x)$. Using $dp_t(x)$ in Lemma 30 and Itô's formula, we have that

$$\begin{aligned}
d\log p_t(x) &= \frac{dp_t(x)}{p_t(x)} - \frac{1}{2}\frac{d[p_t(x)]_t}{p_t(x)^2} \\
&= (x-\mu_t)^\top dW_t - \frac{1}{2}(x-\mu_t)^\top(x-\mu_t)dt \\
&= x^\top(dW_t + \mu_t dt) - \frac{1}{2}\|x\|^2 dt + g(t) \\
&= x^\top dc_t - \frac{1}{2}\|x\|^2 dt + g(t)
\end{aligned}$$

where $g(t)$ represents terms independent of x and the first two terms explain the form of $p_t(x)$ and the appearance of the Gaussian.

Using Itô's formula again, we can compute the change of the covariance matrix:

$$dA_t = \int_{\mathbb{R}^n} (x-\mu_t)(x-\mu_t)^\top \cdot (x-\mu_t)^\top dW_t \cdot p_t(x)dx - A_t^2 dt.$$

3.2.3 Applications

Better KLS bound. Eldan introduced stochastic localization [35] and used it to prove that the thin-shell conjecture implies the KLS conjecture up to a logarithmic factors. We later adapted his idea to get a better bound on KLS constant itself. Since our proof is slightly simpler and more direct than Eldan's proof, we choose to only discuss our proof in full detail. We will discuss the difference between our proof and his proof in the next section.

Theorem 32 ([65]). *For any logconcave density p in $\mathbb{R}^n$ with covariance matrix A,*

$$\psi_p \gtrsim \frac{1}{\left(\mathrm{Tr}\left(A^2\right)\right)^{1/4}}.$$

In particular, we have $\psi_p \gtrsim n^{-\frac{1}{4}}$ for any isotropic logconcave p.

We now outline the proof. For this, we use the simplest choice in stochastic localization, namely a pure random direction chosen from the uniform distribution (i.e., $C_t = I$). The analysis needs a potential function that grows slowly but still maintains good control over the spectral norm of the current covariance matrix. The direct choice of $\|A_t\|_{\mathrm{op}}$, where A_t is the covariance matrix of the distribution at time t, is hard to control. We use the potential $\Phi_t = \mathrm{Tr}(A_t^2)$. Itô's formula shows that this function evolves as follows:

$$\begin{aligned} d\Phi_t =& -2\mathrm{Tr}(A_t^3)dt + \mathbb{E}_{x,y\sim p_t}\left((x-\mu_t)^\top(y-\mu_t)\right)^3 dt + 2\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t(x-\mu_t)(x-\mu_t)^\top dW_t \\ \stackrel{\mathrm{def}}{=}& \delta_t dt + v_t^\top dW_t. \end{aligned} \tag{3.6}$$

The first term can be viewed as a deterministic drift while the second is stochastic with no bias. To bound both terms, we use the following lemmas. The first one below is a folklore reverse Hölder inequality and can be proved using the localization lemma (see e.g., [71]).

Lemma 33 (Logconcave moments). *Given a logconcave density p in $\mathbb{R}^n$, and any positive integer k,*

$$\mathbb{E}_{x\sim p}\|x\|^k \le (2k)^k\left(\mathbb{E}_{x\sim p}\|x\|^2\right)^{k/2}.$$

Using this lemma and the Cauchy-Schwarz inequality, we have the following moment bounds.

Lemma 34. *Given a logconcave distribution p with mean μ and covariance A,*

1. $\mathbb{E}_{x,y\sim p}\left|\langle x-\mu, y-\mu\rangle\right|^3 \lesssim \mathrm{Tr}\left(A^2\right)^{3/2}$.
2. $\left\|\mathbb{E}_{x\sim p}(x-\mu)(x-\mu)^\top A(x-\mu)\right\|_2 \lesssim \|A\|_{\mathrm{op}}^{1/2}\,\mathrm{Tr}\left(A^2\right)$.

Proof. Without loss of generality, we can assume $\mu = 0$.

For the first statement, we fix x and apply Lemma 33 to show that

$$\mathbb{E}_{y\sim p}\left|\langle x,y\rangle\right|^3 \lesssim \left(\mathbb{E}_{y\sim p}\langle x,y\rangle^2\right)^{3/2} = \left(x^\top Ax\right)^{3/2} = \left\|A^{1/2}x\right\|_2^3.$$

Then we note that $A^{1/2}x$ follows a logconcave distribution (Lemma 2) with mean 0 and covariance A^2 and hence Lemma 33 to see that

$$\mathbb{E}_{x\sim p}\left\|A^{1/2}x\right\|_2^3 \lesssim \left(\mathbb{E}_{x\sim p}\left\|A^{1/2}x\right\|^2\right)^{3/2} = \mathrm{Tr}\left(A^2\right)^{3/2}.$$

Therefore, we have that

$$\mathbb{E}_{x,y\sim p}\left|\langle x,y\rangle\right|^{3}\lesssim\mathbb{E}_{x\sim p}\left\Vert A^{1/2}x\right\Vert^{3}\lesssim\mathrm{Tr}\left(A^{2}\right)^{3/2}.$$

For the second statement,

$$\begin{aligned}\left\Vert \mathbb{E}_{x\sim p}x\cdot x^{\top}Ax\right\Vert_{2} &= \max_{\|\zeta\|_{2}\leq1}\mathbb{E}_{x\sim p}x^{\top}\zeta\cdot x^{\top}Ax\\ &\leq \max_{\|\zeta\|_{2}\leq1}\sqrt{\mathbb{E}_{x\sim p}(x^{\top}\zeta)^{2}}\sqrt{\mathbb{E}_{x\sim p}\left(x^{\top}Ax\right)^{2}}\\ &= \|A\|_{\mathrm{op}}^{1/2}\cdot\sqrt{\mathbb{E}_{x\sim p}\left\Vert A^{1/2}x\right\Vert_{2}^{4}}.\end{aligned}$$

For the last term, by a similar argument as before, we can use Lemma 33 shows that

$$\mathbb{E}_{x\sim p}\left\Vert A^{\frac{1}{2}}x\right\Vert_{2}^{4}\lesssim\left(\mathrm{Tr}A^{2}\right)^{2}.$$

This gives the second statement. ◻

The drift term in (3.6) can be bounded using the first inequality in Lemma 34 as

$$\delta_{t}\leq\mathbb{E}_{x,y\sim p_{t}}\left((x-\mu_{t})^{\top}(y-\mu_{t})\right)^{3}\lesssim\mathrm{Tr}\left(A_{t}^{2}\right)^{3/2}=\Phi_{t}^{3/2} \tag{3.7}$$

where we also used that the term $-2\mathrm{Tr}(A_{t}^{3})$ is negative since A_{t} is positive semi-definite. The martingale coefficient v_{t} can be bounded in magnitude using the second inequality:

$$\left\Vert v_{t}\right\Vert_{2}\leq\left\Vert \mathbb{E}_{x\sim p_{t}}(x-\mu_{t})^{\top}A_{t}(x-\mu_{t})(x-\mu_{t})\right\Vert_{2}\leq\left\Vert A_{t}\right\Vert_{\mathrm{op}}^{1/2}\mathrm{Tr}(A_{t}^{2})\lesssim\Phi_{t}^{5/4}.$$

Together we have the simplified expression

$$d\Phi_{t}\lesssim\Phi_{t}^{3/2}dt+\Phi_{t}^{5/4}dW_{t}$$

So the drift term grows roughly as $\Phi^{3/2}t$ while the stochastic term grows as $\Phi_{t}^{5/4}\sqrt{t}$. Thus, both bounds indicate that for t up to $O(\frac{1}{\sqrt{\mathrm{Tr}A^{2}}})$, the potential Φ_{t} remains $O(\mathrm{Tr}A^{2})$, i.e., $\mathrm{Tr}(A_{t}^{2})$ grows only by a constant factor.

We can use this as follows. Fix any subset $E\subset\mathbb{R}^{n}$ of measure $p(E)=\int_{E}p(x)dx=\frac{1}{2}$. We will argue that the set remains nearly balanced for a while. To see this, let $g_{t}=p_{t}(E)$ and note that

$$dg_{t}=\left\langle \int_{E}(x-\mu_{t})p_{t}(x)dx,dW_{t}\right\rangle$$

where the integral might not be 0 because it is over the subset E and not all of $\mathbb{R}^{n}$. Hence,

$$\begin{aligned}d[g_{t}]_{t} &= \left\Vert \int_{E}(x-\mu_{t})p_{t}(x)dx\right\Vert_{2}^{2}dt\\ &= \max_{\|\zeta\|_{2}\leq1}\left(\int_{E}\zeta^{\top}(x-\mu_{t})p_{t}(x)dx\right)^{2}dt\\ &\leq \max_{\|\zeta\|_{2}\leq1}\int_{\mathbb{R}^{n}}\left(\zeta^{\top}(x-\mu_{t})\right)^{2}p_{t}(x)dx\cdot p_{t}(E)dt\\ &\leq \max_{\|\zeta\|_{2}\leq1}\left(\zeta^{\top}A_{t}\zeta\right)dt=\left\Vert A_{t}\right\Vert_{\mathrm{op}}dt.\end{aligned}$$

Thus, g_{t} is bounded by a random process with variance $\|A_{t}\|_{\mathrm{op}}$ at time t. For $0\leq T\lesssim\frac{1}{\sqrt{\mathrm{Tr}A^{2}}}$, the total variance accumulated in the time period $[0,T]$ is

$$\int_{0}^{T}\left\Vert A_{s}\right\Vert_{\mathrm{op}}ds\lesssim\int_{0}^{T}\mathrm{Tr}(A_{s}^{2})^{\frac{1}{2}}ds\lesssim1.$$

Hence, we get that the set E remains bounded in measure between $\frac{1}{4}$ and $\frac{3}{4}$ till time $T = \frac{c}{\sqrt{\mathrm{Tr}A^2}}$ for some universal constant c.

But at this time, the density p_T has a Gaussian component with coefficient T and hence the Cheeger constant is $\Omega(\sqrt{T})$ by Theorem 25. Hence, we have the following:

$$\begin{aligned}
p(\partial E) &= \mathbb{E}p_T(\partial E) \\
&\gtrsim \sqrt{T}\cdot\mathbb{P}(\frac{1}{4}\leq p_T(E)\leq\frac{3}{4}) \\
&\gtrsim \sqrt{T} \\
&\gtrsim (\mathrm{Tr}A^2)^{-\frac{1}{4}}
\end{aligned}$$

where the first equality follows from the fact that p_t is a martingale, the second inequality follows from $\psi_{p_T}\gtrsim\sqrt{T}$ (Theorem 25) and the third inequality follows from the fact the set E remains bounded in measure between $\frac{1}{4}$ and $\frac{3}{4}$ till time $T=\frac{c}{\sqrt{\mathrm{Tr}A^2}}$ with at least constant probability. This completes the proof of Theorem 32.

Reduction from thin shell to KLS. In Eldan's original proof, he used the stochastic process with control matrix $C_t = A_t^{-\frac{1}{2}}$. For this, the change of the covariance matrix is as follows:

$$dA_t = \int_{\mathbb{R}^n}(x-\mu_t)(x-\mu_t)^\top\cdot(x-\mu_t)^\top A_t^{-\frac{1}{2}}dW_t\cdot p_t(x)dx - A_t dt.$$

To get the reduction from thin shell to KLS, we use the potential $\Phi_t = \mathrm{Tr}A_t^q$ with a suitably large integer q for better control of the spectral norm.

$$\begin{aligned}
d\Phi_t =& -q\Phi_t dt + q\mathbb{E}_{x\sim p_t}(x-\mu_t)^\top A_t^{q-1}(x-\mu_t)(x-\mu_t)^\top A_t^{-\frac{1}{2}}dW_t \\
&+\frac{q}{2}\sum_{\alpha+\beta=q-2}\mathbb{E}_{x,y\sim p_t}(x-\mu_t)^\top A_t^{\alpha}(y-\mu_t)(x-\mu_t)^\top A_t^{\beta}(y-\mu_t)(x-\mu_t)^\top A_t^{-1}(y-\mu_t)dt.
\end{aligned}$$

The stochastic term can be bounded using the same proof as in the second inequality of Lemma 34. However, the last term is more complicated. Eldan used the thin shell conjecture to bound the last term and showed that

$$d\Phi_t \lesssim q\Phi_t dW_t + q^2\sigma(n)^2\log n\cdot\Phi_t dt$$

where $\sigma(n)=\sup_p\sigma_p$ is the maximum thin-shell constant over all isotropic logconcave densities p in $\mathbb{R}^n$.

For an isotropic distribution, $\Phi_0 = n$. Hence, we have that $\Phi_t\lesssim n^{O(1)}$ for $0\leq t\leq\frac{1}{q^2\sigma(n)^2}$. By choosing $q=O(\log n)$, we have that $\|A_t\|_{\mathrm{op}}\lesssim 1$ for $0\leq t\leq\frac{1}{\sigma(n)^2\log^2 n}$. By a similar proof as before, this gives that $\psi_p\gtrsim\frac{1}{\sigma(n)\log n}$.

Theorem 35 ([35]). *For any isotropic logconcave density p in $\mathbb{R}^n$, we have $\psi_p\gtrsim\frac{1}{\sigma(n)\log n}$ where $\sigma(n)=\sup_p\sigma_p$ is the maximum thin-shell constant over all isotropic logcave densities p in $\mathbb{R}^n$.*

Tight log-Sobolev constant and improved concentration. One can view the slicing conjecture as being weaker than the thin-shell conjecture and the thin shell conjecture as weaker than the KLS conjecture. Naturally one may ask if there is a conjecture stronger than the KLS conjecture. It is known that KLS conjecture is equivalent to proving Poincaré constant is $\Theta(1)$ for any isotropic logconcave distribution. It is also known that log-Sobolev constant (defined below) is stronger than the Poincaré constant. So, a natural question is whether the log-Sobolev constant is $\Theta(1)$ for any isotropic logconcave distribution? We first remind the reader of the definition.

Definition 36. For any distribution p, we define the the log-Sobolev constant ρ_p be the largest number such that for every smooth f with $\int f^2(x)p(x)dx=1$, we have that

$$\int|\nabla f(x)|^2 p(x)dx\gtrsim\rho_p\int f^2(x)\log f(x)^2\cdot p(x)dx.$$

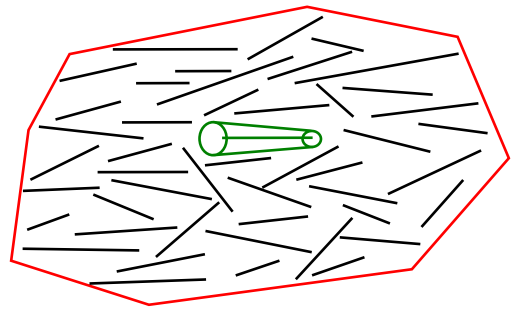


Figure 3.3: Needle decomposition

The result of [48] (Theorem 26) shows that $\rho_p \geq \frac{1}{D^2}$ for any isotropic logconcave measure with support of diameter D. Recently, we proved the following tight bound.

Theorem 37 ([64]). *For any isotropic logconcave density p in $\mathbb{R}^n$ with support of diameter D, the log-Sobolev constant satisfies $\rho_p \gtrsim \frac{1}{D}$. This is the best possible bounds up to a constant.*

The proof uses the same process as Theorem 32 with a different potential function that allows one to get more control on $\|A_t\|_{\mathrm{op}}$. This potential is a Stieltjes-type barrier function defined as follows. Let $u(A_t)$ be the solution to

$$\mathrm{Tr}((uI - A_t)^{-2}) = n \text{ and } A_t \preceq uI \tag{3.8}$$

Note that this is the same as $\sum_{i=1}^{n} \frac{1}{(u-\lambda_i)^2} = n$ and $\lambda_i \leq u$ for all i where λ_i are the eigenvalues of A_t. Such a potential was used to to build graph sparsifiers [14, 4, 61, 62], to understand covariance estimation [83] and to solve bandit problems [9].

The next theorem is a large-deviation inequality based on the same proof technique.

Theorem 38 ([64]). *For any L-Lipschitz function g in $\mathbb{R}^n$ and any isotropic logconcave density p, we have that*

$$\mathbb{P}_{x\sim p}\left(|g(x) - \mathrm{med}_{x\sim p} g(x)| \geq c \cdot L \cdot t\right) \leq \exp\left(-\frac{t^2}{t+\sqrt{n}}\right).$$

Furthermore, the same conclusion holds with $\mathrm{med}_{x\sim p} g(x)$ replaced by $\mathbb{E}_{x\sim p} g(x)$.

For the Euclidean norm $g(x) = \|x\|$, the range $t \geq \sqrt{n}$ is a well-known inequality proved by Paouris [78] and later refined by Guedon and Milman [44] to $\exp(-\min(\frac{t^3}{n}, t))$. The bound above improves and generalizes these bounds.

3.3 Needle decompositions

We describe a "combinatorial" approach to resolving the KLS conjecture. The idea of localization was to reduce an isoperimetric inequality in $\mathbb{R}^n$ to a similar inequality in one dimension, by arguing that if the original inequality were false, there would be a one-dimensional counterexample. Alternatively, one can view localization as an inductive process — the final inequality is a weighted sum of inequalities for each component of a partition into needles, viz. a *needle decomposition*. For this to be valid, the partition should maintain the relative measure of the subset S whose isoperimetry is being considered. To be useful, the Cheeger constant of each needle should be approximately as large as desired. In fact, it suffices if some constant fraction of needles (by measure) in a needle decomposition has good isoperimetry, i.e., small spectral norm, i.e., variance equal to $O(\|A\|_{\mathrm{op}})$.

Definition 39. An ϵ-thin cylinder decomposition of a convex body K is a partition of K by hyperplane cuts so that each part P is contained in a cyclinder whose radius is at most ϵ. The limit of a sequence of needle decompositions with $\epsilon \to 0$ is a needle decomposition with weighting $w(P)$ over the limiting set of needles $\mathbf{P}$.

Theorem 40. *Let* $\mathbf{P}$ *be a needle decomposition of an isotropic convex body* K *by hyperplane cuts s.t. for some* $\int_P f dx = 0$ *for each needle* P *in the decomposition, where* $f : \mathbb{R}^n \to \mathbb{R}$ *is a continuous function. Suppose also that a random needle chosen with density proportional to* $w(P)$ *satisfies* $\mathbb{P}_{P\in\mathbf{P}}(\|A_P\|_{\text{op}} \le C) \ge c$*, for constants* $c, C > 0$*. Then* $\psi_K \gtrsim 1$*.*

The proof of the theorem is simple. We fix a subset $S \subset K$ of measure $a \le 1/2$ and choose f to be the function which maintains the fraction of volume taken up S in each part; thus, the relative measure of S in each needle is a. Next, using one-dimensional isoperimetry, the measure of the boundary of S in each needle P is at least $\Omega(1)\frac{a}{\|A_P\|_{\text{op}}^{1/2}}w(P)$. This is $\Omega(a)w(P)$ for at least c fraction of the needles (by their weight) by the assumption of the theorem. Hence

$$\frac{\text{vol}_{n-1}(\partial S)}{\text{vol}(S)} \ge c \cdot \Omega(a) = \Omega(a)$$

showing that $\psi_K = \Omega(1)$.

This puts the focus on whether there exist needle decompositions that have bounded operator norm for some constant measure of the needles. This approach was used in [28] to bound the isoperimetry of star-shaped bodies.

As far as we know, the bounded operator norm property might be true for *any* needle decomposition! We conclude with this as a question:

Let $\mathbf{P}$ be a partition of an isotropic logconcave density p in $\mathbb{R}^n$ by hyperplanes. Is it true that there always exists a subset of parts $Q \subset \mathbb{P}$ such that (1) $p(Q) \ge c$ and (2) $\text{Var}(P) \le C$ for each $P \in Q$? (c, C are absolute constants, and $\text{Var}(P)$ is the variance of a random point from the part P drawn with density proportional to p).

4 Open Problems

Here we discuss some intriguing open questions related to the KLS conjecture (besides resolving it!), asymptotic convex geometry and efficient algorithms in general.

Deterministic volume. Efficient algorithms for volume computation are randomized, and this is unavoidable if access to the input body is only through an oracle [37, 13]. However, for explicit polytopes given as $Ax \ge b$, the only known hardness is for exact computation [33] and it is possible that there is a deterministic polynomial-time approximation scheme. Such an algorithm would be implied if P=BPP. Thus, finding a deterministic polynomial-time algorithm for estimating the volume of a polytope to within a factor of 2 (say) is an outstanding open problem.

Lower bound for sampling. The complexity of sampling an isotropic logconcave density in $\mathbb{R}^n$, assuming the KLS conjecture is $O^*(n^2)$ from a warm start. Is this the best possible? Can we show an $\Omega(n^2)$ lower bound for randomized algorithms?

Faster sampling and isotropy. The current bottleneck for faster sampling (in the general oracle model) is the complexity of isotropic transformation, which is currently $O^*(n^4)$ [70]. Cousins and Vempala have conjectured that the following algorithm will terminate in $O(\log n)$ iterations and produce a nearly-istropic body. Each iteration needs $O^*(n)$ samples and takes $O^*(n^2) \times O^*(n) = O^*(n^3)$ steps.

Repeat:

1. Estimate the covariance of the standard Gaussian density restricted to the current convex body.

2. If the covariance has eigenvalues smaller than some constant, apply a transformation to make this distribution isotropic.

Another avenue for improvement is in the number of arithmetic operations per oracle query. This is $\widetilde{O}(n^2)$ for all the oracle-based methods since each step must deal with a linear transformation. A random process that could be faster is Coordinate Hit-and-Run. In this, a coordinate basis is fixed, and at each step, we pick one of the coordinate directions at random, compute the chord through the current point in that direction and go to a random point along the chord. It is open to show that the mixing time/conductance of this process is polynomial and perhaps of the same order as Hit-and-Run, thus potentially a factor of n faster overall.

Manifold KLS. In [63], we proved the following theorem, motivated by the goal of faster sampling and volume computation of polytopes.

Lemma 41. *Let $\phi : K \subset \mathbb{R}^n \to \mathbb{R}$ be a convex function defined over a convex body K such that $D^4\phi(x)[h,h,h,h] \geq 0$ for all $x \in K$ and $h \in \mathbb{R}^n$. Given any partition S_1, S_2, S_3 of K with $d = \min_{x\in S_1, y\in S_2} d(x,y)$, i.e., the minimum distance between S_1 and S_2 in the Riemannian metric induced by the Hessian of ϕ. For any $\alpha > 0$,*

$$\frac{\int_{S_3} e^{-\alpha\phi(x)}dx}{\min\left\{\int_{S_1} e^{-\alpha\phi(x)}dx, \int_{S_2} e^{-\alpha\phi(x)}\,dx\right\}} \gtrsim \sqrt{\alpha}\cdot d.$$

The special case when $\phi(x) = \|x\|^2$ and d is the Euclidean metric is given by Theorem 25. What are interesting generalizations of this theorem and the KLS conjecture to the manifold setting?

Acknowledgement. This work was supported in part by NSF Awards CCF-1563838, CCF-1717349 and CCF-1740551. We are grateful to Ronen Eldan, Bo'az Klartag and Emanuel Milman for helpful discussions and Yan Kit Chim for making the illustrations.

References

[1] Jacob Abernethy and Elad Hazan. Faster convex optimization: Simulated annealing with an efficient universal barrier. In *International Conference on Machine Learning*, pages 2520–2528, 2016.

[2] R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, 23:535–561, 2010.

[3] Radosław Adamczak, Alexander E Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Sharp bounds on the rate of convergence of the empirical covariance matrix. *Comptes Rendus Mathematique*, 349(3):195–200, 2011.

[4] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 237–245. ACM, 2015.

[5] David Alonso-Gutiérrez and Jesús Bastero. *Approaching the Kannan-Lovász-Simonovits and variance conjectures*, volume 2131. Springer, 2015.

[6] D. Applegate and R. Kannan. Sampling and integration of near log-concave functions. In *STOC*, pages 156–163, 1991.

[7] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D Milman. *Asymptotic Geometric Analysis, Part I*, volume 202. 2015.

[8] Mark Ashbaugh and Rafael Benguria. The problem of queen dido, 2010.

[9] Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2013.

[10] K. Ball. Lecture. *Institut Henri Poincare, Paris*, 2006.

[11] Keith Ball. Logarithmically concave functions and sections of convex sets in rn. *Studia Math*, 88(1):69–84, 1988.

[12] Keith Ball and Van Huang Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Mathematica*, 213(1):81–96, 2012.

[13] I. Bárány and Z. Füredi. Computing the volume is difficult. *Discrete Comput. Geom.*, 2(4):319–326, 1987.

[14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.

[15] Alexandre Belloni, Tengyuan Liang, Hariharan Narayanan, and Alexander Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In *Conference on Learning Theory*, pages 240–265, 2015.

[16] D. Bertsimas and S. Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, 2004.

[17] S. G. Bobkov and A. Koldobsky. *On the Central Limit Property of Convex Bodies*, pages 44–52. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[18] Sergey Bobkov and Mokshay Madiman. Entropy and the hyperplane conjecture in convex geometry. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1438–1442. IEEE, 2010.

[19] SG Bobkov. On concentration of distributions of random weighted sums. *Annals of probability*, pages 195–215, 2003.

[20] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.

[21] J Bourgain, B Klartag, and V Milman. Symmetrization and isotropic constants of convex bodies. In *Geometric aspects of functional analysis*, pages 101–115. Springer, 2004.

[22] Jean Bourgain. On high dimensional maximal functions associated to convex bodies. *American Journal of Mathematics*, 108(6):1467–1476, 1986.

[23] Silouanos Brazitikos, Apostolos Giannopoulos, Petros Valettas, and Beatrice-Helen Vritsiou. *Geometry of isotropic convex bodies*, volume 196. American Mathematical Society Providence, 2014.

[24] Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. *arXiv preprint arXiv:1412.1587*, 2014.

[25] Sébastien Bubeck, Yin Tat Lee, and Ronen Eldan. Kernel-based methods for bandit convex optimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 72–85. ACM, 2017.

[26] Peter Buser. A note on the isoperimetric constant. In *Annales scientifiques de l'École Normale Supérieure*, volume 15, pages 213–230. Elsevier, 1982.

[27] Anthony Carbery and James Wright. Distributional and l^ q norm inequalities for polynomials over convex bodies in r^ n. *Mathematical research letters*, 8(3):233–248, 2001.

[28] K. Chandrasekaran, D. Dadush, and S. Vempala. Thin partitions: Isoperimetric inequalities and a sampling algorithm for star shaped bodies. In *SODA*, pages 1630–1645, 2010.

[29] Jeff Cheeger. *A lower bound for the smallest eigenvalue of the Laplacian*, pages 195–199. Princeton Univ. Press, 1969.

[30] B. Cousins and S. Vempala. A cubic algorithm for computing Gaussian volume. In *SODA*, pages 1215–1228, 2014.

[31] B. Cousins and S. Vempala. Bypassing KLS: Gaussian cooling and an $O^*(n^3)$ volume algorithm. In *STOC*, pages 539–548, 2015.

[32] Nikos Dafnis and Grigoris Paouris. Small ball probability estimates, $\psi 2$-behavior and the hyperplane conjecture. *Journal of Functional Analysis*, 258(6):1933–1964, 2010.

[33] M. E. Dyer and A. M. Frieze. Computing the volume of a convex body: a case where randomness provably helps. In *Proc. of AMS Symposium on Probabilistic Combinatorics and Its Applications*, pages 123–170, 1991.

[34] M. E. Dyer, A. M. Frieze, and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. In *STOC*, pages 375–381, 1989.

[35] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23:532–569, 2013.

[36] R. Eldan and B. Klartag. Approximately gaussian marginals and the hyperplane conjecture. *Contermporary Mathematics*, 545, 2011.

[37] G. Elekes. A geometric inequality and the complexity of computing volume. *Discrete & Computational Geometry*, pages 289–292, 1986.

[38] Vitaly Feldman, Cristobal Guzman, and Santosh Vempala. Statistical query algorithms for mean vector estimation and stochastic convex optimization. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 1265–1277, 2017.

[39] B. Fleury. Concentration in a thin Euclidean shell for log-concave measures. *J. Funct. Anal.*, 259(4):832–841, 2010.

[40] Bruno Fleury, Olivier Guédon, and Grigoris Paouris. A stability result for mean width of lp-centroid bodies. *Advances in Mathematics*, 214(2):865–877, 2007.

[41] Matthieu Fradelizi and Olivier Guédon. The extreme points of subsets of s-concave probabilities and a geometric localization theorem. *Discrete & Computational Geometry*, 31(2):327–335, 2004.

[42] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983.

[43] B. Grunbaum. Partitions of mass-distributions and convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261, 1960.

[44] Olivier Guedon and Emanuel Milman. Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geometric and Functional Analysis*, 21(5):1043–1068, 2011.

[45] A. T. Kalai and S. Vempala. Simulated annealing for convex optimization. *Math. Oper. Res.*, 31(2):253–266, 2006.

[46] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

[47] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms*, 11:1–50, 1997.

[48] Ravi Kannan, László Lovász, and Ravi Montenegro. Blocking conductance and mixing in random walks. *Combinatorics, Probability & Computing*, 15(4):541–570, 2006.

[49] Alexander Karzanov and Leonid Khachiyan. On the conductance of order Markov chains. *Order*, 8(1):7–15, 1991.

[50] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. and Funct. Anal.*, 16(6):1274–1290, 2006.

[51] B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168:91–131, 2007.

[52] B. Klartag. Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.*, 245:284–310, 2007.

[53] B. Klartag. Needle decompositions in riemannian geometry. 2017.

[54] B. Klartag and E. Milman. Centroid bodies and the logarithmic laplace transform–a unified approach. *Journal of Functional Analysis*, 262(1):10–34, 2012.

[55] Michel Ledoux. A simple analytic proof of an inequality by p. buser. *Proceedings of the American mathematical society*, 121(3):951–959, 1994.

[56] Michel Ledoux. Concentration of measure and logarithmic sobolev inequalities. *LECTURE NOTES IN MATHEMATICS-SPRINGER VERLAG-*, pages 120–216, 1999.

[57] Michel Ledoux. Concentration of measure and logarithmic sobolev inequalities. *Seminaire de probabilites de Strasbourg*, 33:120–216, 1999.

[58] Yin Tat Lee and Aaron Sidford. Path finding methods for linear programming: Solving linear programs in õ (vrank) iterations and faster algorithms for maximum flow. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 424–433. IEEE, 2014.

[59] Yin Tat Lee and Aaron Sidford. Efficient inverse maintenance and faster algorithms for linear programming. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 230–249. IEEE, 2015.

[60] Yin Tat Lee, Aaron Sidford, and Sam Chiu-wai Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1049–1065. IEEE, 2015.

[61] Yin Tat Lee and He Sun. Constructing linear-sized spectral sparsification in almost-linear time. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 250–269. IEEE, 2015.

[62] Yin Tat Lee and He Sun. An sdp-based algorithm for linear-sized spectral sparsification. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 678–687. ACM, 2017.

[63] Yin Tat Lee and Santosh S Vempala. Convergence rate of riemannian hamiltonian monte carlo and faster polytope volume computation. *arXiv preprint arXiv:1710.06261*, 2017.

[64] Yin Tat Lee and Santosh S Vempala. Stochastic localization+ stieltjes barrier= tight bound for log-sobolev. *arXiv preprint arXiv:1712.01791*, 2017.

[65] Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *Proc. of IEEE FOCS*, 2017.

[66] Peter Li and Shing Tung Yau. Estimates of eigenvalues of a compact riemannian manifold. *Geometry of the Laplace operator*, 36:205–239, 1980.

[67] L. Lovász. How to compute the volume? *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, pages 138–151, 1990.

[68] L. Lovász and M. Simonovits. Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *ROCS*, pages 482–491, 1990.

[69] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. In *Random Structures and Alg.*, volume 4, pages 359–412, 1993.

[70] L. Lovász and S. Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. Syst. Sci.*, 72(2):392–417, 2006.

[71] L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, 2007.

[72] Irini Perissinaki Milla Anttila, Keith Ball. The central limit problem for convex bodies. *Transactions of the American Mathematical Society*, 355(12):4723–4735, 2003.

[73] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1):1–43, 2009.

[74] Vitali D Milman and Alain Pajor. Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. In *Geometric aspects of functional analysis*, pages 64–104. Springer, 1989.

[75] Hariharan Narayanan and Alexander Rakhlin. Random walk approach to regret minimization. In *Advances in Neural Information Processing Systems*, pages 1777–1785, 2010.

[76] Yu. Nesterov and A. Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. SIAM, 1994.

[77] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[78] G. Paouris. Concentration of mass on convex bodies. *Geometric and Functional Analysis*, 16:1021–1049, 2006.

[79] Grigoris Paouris. Small ball probability estimates for log-concave measures. *Transactions of the American Mathematical Society*, 364(1):287–308, 2012.

[80] Lawrence E Payne and Hans F Weinberger. An optimal poincaré inequality for convex domains. *Archive for Rational Mechanics and Analysis*, 5(1):286–292, 1960.

[81] Miklós Simonovits. How to compute the volume in high dimension? *Math. Program.*, 97(1-2):337–374, 2003.

[82] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82:93–133, 1989.

[83] Nikhil Srivastava, Roman Vershynin, et al. Covariance estimation for distributions with 2+eps moments. *The Annals of Probability*, 41(5):3081–3111, 2013.

[84] P Stavrakakis and P Valettas. On the geometry of log-concave probability measures with bounded log-sobolev constant. In *Asymptotic Geometric Analysis*, pages 359–380. Springer, 2013.

[85] Jacob Steiner. Einfache beweise der isoperimetrischen hauptsätze. *Journal für die reine und angewandte Mathematik*, 18:281–296, 1838.

[86] Vladimir Nikolaevich Sudakov and Boris S Tsirel'son. Extremal properties of semi-spaces for spherically symmetrical measures. *Zapiski Nauchnykh Seminarov POMI*, 41:14–24, 1974.

[87] S. Vempala. Geometric random walks: A survey. *MSRI Combinatorial and Computational Geometry*, 52:573–612, 2005.

[88] V.G.Maz'ja. Classes of domains and imbedding theorems for function spaces. *Dokl. Acad. Nauk SSSR (Engl. transl. Soviet Math. Dokl., 1 (1961) 882-885)*, 3:527–530, 1960.