On Yuansi Chen's work on the KLS conjecture

Lectures by Bo'az Klartag*

Winter School at the Hausdorff Institute, January 2021

Lecture 1 – Background on the KLS conjecture

The KLS conjecture is a question regarding the isoperimetric inequality in high-dimensional convex bodies, which has several consequences in high-dimensional geometry. Six and a half weeks ago, Yuansi Chen uploaded a preprint to the arXiv which represents significant progress towards this conjecture. These three lectures are devoted to this progress, and to the many ideas upon which it builds.

Isoperimetry: Let (X, d, μ) be a measure-metric space, i.e., (X, d) is a metric space and μ a Borel measure on X. Each measurable subset A is associated with its measure $\mu(A)$, as well as with its boundary measure

$$\mu^+(\partial A) = \liminf_{\varepsilon \to 0^+} \frac{\mu(A+\varepsilon) - \mu(A)}{\varepsilon}$$

where $A + \varepsilon = \{x \in X \, ; \, \inf_{y \in A} d(x,y) < \varepsilon\}$ is the ε-neighborhood of A.

Example: Assume that $X = K \subseteq \mathbb{R}^n$ is an open set with smooth boundary (Lipschitz is enough). Now d is the Euclidean metric in $\mathbb{R}^n$. The measure μ has a density $p : K \to [0, \infty)$, continuous up to the boundary. Then for a domain $A \subseteq \mathbb{R}^n$ with smooth boundary,

$$\mu^+(\partial A) = \int_{K \cap \partial A} p.$$

Isoperimetric problem in (X, d, μ)**:** For a fixed $t > 0$, among all $A \subseteq X$ of measure t, minimize the boundary measure $\mu^+(\partial A)$.

Exact solution: In $\mathbb{R}^n, S^n, \mathbb{H}^n$, half-space, B^n. The hamming cube $\{0,1\}^n$ if we slightly change the definition of the boundary measure, $\varepsilon = 1$ in place of a limit. An example which is relevant to us is

$$(\mathbb{R}^n, d_{Euclid}, \gamma_n)$$

*Videos should be available at https://www.him.uni-bonn.de/index.php?id=4133 where γ_n is the standard Gaussian measure in $\mathbb{R}^n$. These are highly symmetric or highly structured situations.

Another case in which an exact solution may be described: If $K \subseteq \mathbb{R}^2$ is convex, μ the Lebesgue measure on K, and d is Euclidean – Then the boundary of the extremal sets is either a single circular arc or a straight line, and in both cases the intersection with the boundary is orthogonal.

Non-trivial approximate solutions: There are examples in combinatorics ("expander graphs"), in geometric group theory and in Riemannian geometry. In these lectures we are interested in *convexity in Euclidean space.* Throughout these lectures, from now on we assume that

$$X = K \subseteq \mathbb{R}^n$$

is an open, convex set, and d is the Euclidean metric. What about the measure μ? If K is bounded, then we may take μ as the uniform, Lebesgue measure on K normalized to a probability measure. More generally: We assume that the probability measure μ has a log-concave density in $\mathbb{R}^n$. That is, the density $p : \mathbb{R}^n \to [0, \infty)$ satisfies

$$p(\lambda x + (1-\lambda)y) \geq p(x)^\lambda p(y)^{1-\lambda} \qquad \forall x, y \in \mathbb{R}^n, 0 < \lambda < 1.$$

This includes the uniform measure on convex sets, the Gaussian measures, and it is a rather stable condition: Closed under convolutions, products, weak limits etc. If p is smooth and positive, and $p = e^{-\rho}$, then log-concavity is equivalent to

$$\nabla^2 \rho \geq 0.$$

Is convexity relevant to isoperimetry? Of course. There are quite a few results in this direction, essentially going back to the Poincaré inequality from the 19^{th} century, "spectral gap under convexity assumption". We will discuss it soon. In the meanwhile let us mention

Theorem (Sternberg-Zumbrun '99 –convex body case, E. Milman '09 – generalization to log–concave)**.** *The isoperimetric profile I is concave, where*

$$I(t) = \inf_{A \subseteq \mathbb{R}^n} \left\{ \mu^+(\partial A) \, ; \, \mu(A) = t \right\}.$$

It is also symmetric about $1/2$.

This is a difficult theorem, because it relies on regularity of the minimizer and on results in geometric measure theory.

Corollary. *("Enough to partition into two halves") We have*

$$I(t) \geq 2I\left(\frac{1}{2}\right)\min\{t, 1-t\}.$$

Hence for any measurable $A \subseteq \mathbb{R}^n$,

$$\mu^+(\partial A) \geq 2I\left(\frac{1}{2}\right)\min\{\mu(A), \mu(\mathbb{R}^n \setminus A)\}.$$

Definition. *The* *Isoperimetric Constant* *(or Cheeger constant)* $\psi = \psi_\mu$ *is*

$$\psi = \psi_\mu = \inf_{A \subseteq \mathbb{R}^n} \frac{\mu^+(\partial A)}{\min\{\mu(A), 1-\mu(A)\}}.$$

Attained when $\mu(A) = 1/2$.

How should we partition a convex body into two pieces with small interface?

KLS Conjecture (Kannan-Lovász-Simonovits '95, "Up to constant, bisect by a hyperplane")**.** *There exists a universal constant* $c > 0$ *such that* $\forall \mu$ *in* $\mathbb{R}^n$*, log-concave probability measure,*

$$\psi_\mu \geq c \cdot \inf_{\substack{H \subseteq \mathbb{R}^n \\ \text{halfspace}}} \frac{\mu^+(\partial H)}{\min\{\mu(H), 1-\mu(H)\}}, \tag{1}$$

where the infimum runs over all half-spaces $H \subseteq \mathbb{R}^n$.

Remarks.

1. The infimum over half-spaces is much easier to compute, using the covariance matrix of μ which is $Cov(\mu) = (C_{ij})_{i,j=1,\dots,n}$ where

$$C_{ij} = \int_{\mathbb{R}^n} x_i x_j d\mu(x) - \int_{\mathbb{R}^n} x_i d\mu(x) \int_{\mathbb{R}^n} x_j d\mu(x).$$

   The RHS of (1) is equivalent, up to a universal constant, to $1/\sqrt{\|Cov(\mu)\|_{op}}$. A completely equivalent formulation of the KLS conjecture is that

$$\psi_\mu \geq \frac{c}{\sqrt{\|Cov(\mu)\|_{op}}}.$$

   Why? By Hensley '80, Fradelizi '99: For any $\theta \in S^{n-1}$,

$$\frac{1}{5} \leq \sqrt{Cov(\mu)\theta \cdot \theta} \cdot \sup_{t \in \mathbb{R}} \mu^+(\{x \cdot \theta = t\}) \leq 5.$$

   Thanks to the Prékopa-Leindler inequality (a variant of Brunn-Minkowski), we reduce matters to the following one-dimensional inequality: For any log-concave probability density $p : \mathbb{R} \to [0, \infty)$ with $\int_{-\infty}^{\infty} x p(x) dx = 0$,

$$\frac{1}{5} \leq \sqrt{\int_{-\infty}^{\infty} x^2 p(x) dx} \cdot \sup_{x \in \mathbb{R}} p(x) \leq 5.$$

2. If $Cov(\mu) = \mathrm{Id}$ then we say that μ is isotropic or normalized. The KLS conjecture is equivalent to the bound $\psi_\mu \geq c$ for an isotropic, log-concave probability measure μ.

Denote $\psi_n = \inf_{\mu \text{ in } \mathbb{R}^n} \psi_\mu$, where the infimum runs over all log-concave, isotropic measures in $\mathbb{R}^n$.

Yuansi Chen thm (arXiv, 27/11/20): For any $\varepsilon > 0$ there exists $c_\varepsilon > 0$ such that

$$\psi_n \geq c_\varepsilon / n^\varepsilon.$$

Optimizing over ε we get $\psi_n \geq C \exp(-\sqrt{\log n \log \log n})$.

Earlier bounds:

1. Using needle decomposition/localization: $\psi_n \gtrsim n^{-0.5}$ by KLS '95, improved by Bobkov '07 to $\psi_n \gtrsim n^{-0.25}/\sqrt{\sigma_n}$ for a certain "thin shell parameter" σ_n.
2. Using non-trivial bounds for "thin shell": $\psi_n \gtrsim n^{-0.46}$ by K. '07, Fleury '09 gave $n^{-7/16}$, Guédon and E. Milman '11 gave $n^{-5/12}$.
3. Using Stochastic localization: Eldan '13 gave $n^{-1/3}/\log n$ and Lee-Vempala '17 gave $n^{-1/4}$.

At the time some mathematicians speculated that the Lee-Vempala bound could have been the optimal answer, now we know that it is not.

Applications.

1. **Bourgain's slicing problem.** Given $K \subseteq \mathbb{R}^n$ convex, volume one. Does there exist a hyperplane $H \subseteq \mathbb{R}^n$ with $Vol_{n-1}(K \cap H) \geq c$?

   Still open. We may replace c by $1/L_n$, where: Bourgain '91, $L_n \leq C n^{1/4} \log n$, and K. '06: $L_n \leq C n^{1/4}$.

   Eldan-Klartag '11 proved $L_n \leq C/\psi_n$ (and Ball-Nguyen '13 proved $L_n \lesssim \exp(c/\psi_n^2)$), so now we know
   $$L_n \leq c_\varepsilon n^\varepsilon \qquad \forall \varepsilon > 0.$$

2. **Approx. Gauss marginals.** Let X be a random vector in $\mathbb{R}^n$, log-concave density. Assume that it is normalized (or isotropic) so $\mathbb{E}X = 0$ and $\mathrm{Cov}(X) = \mathrm{Id}$.

   Convex CLT (K., '07) *For most $\theta \in S^{n-1}$ (in sense of measure),*
   $$d\left(\langle X, \theta \rangle, Z\right) \leq C/n^\alpha$$
   for some universal constants $C, \alpha > 0$. Here $Z \sim N(0,1)$ is a standard Gaussian and $d(Y, Z) = \sup_{t \in \mathbb{R}} |\mathbb{P}(Y \leq t) - P(Z \leq t)|$ is the Kolmogorov distance.

Theorem (Bobkov-Chistyakov-Götze '19, plus Chen's bound). *For any $\varepsilon > 0$, can take $\alpha = 1 - \varepsilon$.*

Note that this is a better rate than the Berry-Esseen rate for $\theta = (1, \dots, 1)/\sqrt{n}$.

3. **Cheeger's inequality.** We always have

$$\lambda_\mu \geq \frac{\psi_\mu^2}{4}$$

where $\lambda = \lambda_\mu$ is the Poincaré constant, the "spectral gap", the best constant in the Poincaré inequality: For any smooth $f : \mathbb{R}^n \to \mathbb{R}$,

$$\lambda \cdot Var_\mu(f) \leq \int_{\mathbb{R}^n} |\nabla f|^2 d\mu$$

where $Var_\mu(f) = \int f^2 d\mu - (\int f d\mu)^2$. The Poincaré constant governs the mixing time of heat flow and certain random walks on convex bodies. This is used in algorithms for sampling and for computing the volume of convex bodies. Chen's result improved the complexity bound for such algorithms.

How can we prove isoperimetric inequalities under convexity assumptions? With uniform bounds on log-concavity, things are easier:

Theorem ("More log-concave than Gaussian", Bakry-Ledoux '96). *Suppose that $d\mu/dx = e^{-\rho}$ and $t > 0$ satisfy $\nabla^2 \rho(x) \geq t \cdot \mathrm{Id}$ for all $x \in \mathbb{R}^n$, in the sense of symmetric matrices. Then,*

$$\psi_\mu \geq \sqrt{\frac{2}{\pi}} \cdot \sqrt{t}$$

which is what you get for the Gaussian $e^{-t|x|^2/2}$. Similarly, for the spectral gap, we have $\lambda_\mu \geq t$.

There are a number of ways to prove this: First, Bakry-Emery Γ_2 calculus or Bochner-type formulas. Second, mass transport, or you can use the Caffarelli '00 result, that there exists a contraction transporting the Gaussian to μ. An isoperimetric coefficient is improved by measure-transporting contractions. Another method is:

Convex localization, decomposition of measure into needles (Payne-Weinberger '60, Gromov-Milman '87, LS '93 and KLS '95). Roughly, given $A \subseteq \mathbb{R}^n$ with $\mu(A) = 1/2$, bisect by a hyperplane H so that

$$\frac{\mu(A \cap H^+)}{\mu(H^+)} = \frac{1}{2}, \qquad \frac{\mu(A \cap H^-)}{\mu(H^-)} = \frac{1}{2}$$

where $H^\pm$ are the two half-spaces determined by H. Then bisect again and again, until in the limit you get something lower dimensional, and then induct on the dimension. In the next lecture we will discuss another method, Eldan's stochastic localization, which is motivated by convex localization.

Lecture 2 – Eldan's stochastic localization, Lee-Vempala version

Convex localization involves sharp bisections by hyperplanes: We replace K by $K \cap H^+$ and $K \cap H^-$. In his Ph.D. Eldan worked on an extension from two aspects:

- **Random localization:** Choose hyperplanes randomly, through barycenter.
- **Continuous localization:** Don't cut sharply. Instead, multiply by an affine functional that is very close to the constant function 1.

Let μ be a probability measure with a log-concave density p. For a small $\varepsilon > 0$, in place of cutting into H^+ and H^-, we look at the two measures

$$(1 \pm \varepsilon\langle x - b_\mu, \theta\rangle)p(x),$$

where $b_\mu = \int x d\mu(x)$ is the barycenter. Then,

- We get two measures whose average is the original measure.
- Moreover, assume that $\varepsilon > 0$ is small and μ compactly-supported. Say that Z_t is uniform on $\sqrt{n}S^{n-1}$. Then each of these two measures is still a log-concave probability density! This procedure has the virtues of convex localization.
- Why multiply by an affine function of all functions? With other functions, it is hard to arrange that both measures are log-concave.

Let us now repeat this "soft bisection" process again and again, with θ being random.

<u>**A stochastic process (discrete time)**</u>: Fix a small $\varepsilon = \Delta t > 0$, which will tend to zero in a few moments. Set $p_0(x) = p(x)$ and

$$p_{t+\Delta t}(x) = (1 + \langle x - b_t, \sqrt{\Delta t} Z_t\rangle)p(x)$$

where Z_t for $t = 0, \varepsilon, 2\varepsilon, \ldots$ is a sequence of i.i.d. standard random vectors in $\mathbb{R}^n$, maybe uniform on $\sqrt{n}S^{n-1}$ or maybe standard Gaussians. [The difference is miniscule when ε is small, but let us consider the sphere for now] Here b_t is the barycenter of p_t. We can write it as a difference equation

$$p_{t+\Delta t}(x) = p_t(x) + \langle x - b_t, \sqrt{\Delta t} Z_t\rangle p(x).$$

Then we get a random log-concave probability measure p_t. Since the distribution of $Z - t$ is symmetric, for all $x \in \mathbb{R}^n$,

$$\mathbb{E} p_t(x) = p_0(x).$$

That is, we have a decomposition of the measure p into log-concave pieces.

More precisely, let $\mathcal{F}_t$ be the σ-algebra generated by $(Z_s)_{0\le s\le t}$, all events that are determined up to time t. Then,

$$\mathbb{E}\left(p_{t+\Delta t}|\mathcal{F}_t\right) = p_t(x).$$

This is called a **martingale condition**: The expectation of the next value conditioned on the past, is the present value.

[If Z_t is Gaussian, then there is a small probability that p_t would become negative because the affine functional could become negative in the support. However, this small probability goes to zero as Δt goes to zero. The two possibilities for Z_t yield very similar processes for small Δt.]

If Z_t are i.i.d Gaussian then we could have taken

$$\sqrt{\Delta t} Z_t = W_{t+\Delta t} - W_t =: \Delta W_t,$$

where $(W_t)_{t\ge 0}$ is a standard Brownian motion in $\mathbb{R}^n$, which we fix once and for all.

By moving from discrete time to continuous time we gain precise formulas for derivatives and for changing variables. Thus, by letting $\Delta t \to 0$ we obtain:

<u>**Stochastic Localization**</u>: A stochastic differential equation, a limit of the stochastic difference equation discussed earlier:

$$dp_t(x) = \langle x - b_t, dW_t\rangle p_t(x)$$

where p_0 is given, and $b_t = \int_{\mathbb{R}^n} x p_t(x) dx$.

Theorem. *Existence and Uniqueness of a solution for* $t \in (0,\infty)$. *Moreover, almost surely, for all* t,

- p_t *is a continuous probability density, positive on* $Supp(p) = Supp(p_0)$.
  (This may be proven by considering the evolution equation of $\log p_t$, *written in (2) below)*

- *For any fixed* $x \in \mathbb{R}^n$, *the process* $(p_t(x))_{t\ge 0}$ *is a martingale, i.e.,*

$$\mathbb{E}(p_t(x)|\mathcal{F}_s) = p_s(x) \qquad \textit{for all } t > s.$$

  Here, $\mathcal{F}_t$ *is the* σ*-algebra generated by* $(W_r)_{0\le r\le t}$, *all events determined up to time* t.

Remarks.

1. All stochastic processes $(X_t)_{t\ge 0}$ that we will discuss from now on will be *adapted*, the value of X_t is determined by the path of the Brownian motion up to time t. Our processes cannot see the future.

2. Suppose that $(X_t)_{t\geq 0}$ is an adapted stochastic process, real-valued, with the stochastic differential equation
$$dX_t = \langle v_t, dW_t\rangle + \delta_t dt.$$
What is the meaning of this equation? Formally, through Itô integral. Intuitively, this means that the increment in X_t in an infinitesimal time interval dt is the sum of two terms, one is just $\delta_t dt$ like in ordinary differential equations, and the other is a little new Gaussian vector times v_t. In small scales, the other term is roughly of the order of $\sqrt{dt}$, so it is larger than the $\delta_t dt$ term. Indeed, recall that $\Delta W_t = W_{t+\Delta t} - W_t \overset{(d)}{=} \sqrt{\Delta t} Z$.

3. If δ_t is zero, then X_t is a martingale, the expected increment is zero. Moreover, in general,
$$\frac{d}{dt}\mathbb{E}X_t = \mathbb{E}\delta_t.$$

4. Itô's change-of-variables formula states that for a reasonable function $\varphi : \mathbb{R} \to \mathbb{R}$,
$$d\varphi(X_t) = \varphi'(X_t)dX_t + \frac{1}{2}\varphi''(X_t)|v_t|^2 dt.$$
There are two summands in the Itô formula. The first one involves the first derivative of φ is the familiar term from ordinary differential equations (ODE). The second term – the "Itô term" – involves second derivative of φ. This is roughly because the Brownian motion jumps to distance $\sqrt{dt}$, so the second term in the Taylor expansion would contribute something of the order of dt.

To summarize, for any set $E \subseteq \mathbb{R}^n$ (say with a smooth boundary),
$$p_0(E) = \mathbb{E}p_t(E)$$
where by abuse of notation p_t denotes both the density and the measure, and
$$p_0^+(\partial E) = \mathbb{E}p_t^+(\partial E).$$

The discrete analog of the process was log-concave. In the continuous world we have a much more precise statement. This is one of the virtues of stochastic localization, due to Lee and Vempala:

Proposition. *("More log-concave than the Gaussian") Write $p_t = e^{-\rho_t}$. Then with probability one, for any $x \in Supp(p_0)$,*
$$\nabla^2\rho_t(x) = \nabla^2\rho_0(x) + t \cdot \mathrm{Id} \geq t \cdot \mathrm{Id}.$$
In particular p_t is always more log-concave than $e^{-t|x|^2/2}$.

For the proof, use Itô's formula:

$$dp_t(x) = \langle x - b_t, dW_t \rangle p_t(x)$$

so

$$d \log p_t(x) = \langle x - b_t, dW_t \rangle - \frac{1}{2}|x - b_t|^2 dt = -\frac{|x|^2}{2} dt + \text{affine function of } x. \tag{2}$$

Consequently, the Cheeger constant of p_t is always at least $c\sqrt{t}$. I.e., for any $E \subseteq \mathbb{R}^n$,

$$p_t^+(\partial E) \geq c\sqrt{t} \cdot p_t(E) \cdot (1 - p_t(E))$$

where we abbreviate $p_t(E) = \int_E p_t$, i.e., both the probability measure and its density are denoted by p_t.

Proof strategy of Chen's theorem: Begin with an isotropic, log-concave density $p = p_0$ on $\mathbb{R}^n$. For simplicity let us assume that

$$Supp(p) \subseteq B(0, n^5).$$

This holds automatically when p is uniform on a convex body, and in the general log-concave case there are stability estimates following from Milman/Sternberg-Zumbrun allowing to throw away an exponentially small amount of mass. [Exercise: Show that by truncating to the ball we remove less than Ce^{-cn^2} of the mass. Paouris '06: Less than e^{-cn^5}].

We run the Eldan stochastic localization process, obtain a stochastic process of log-concave measures $(p_t)_{t \geq 0}$. Set

$$A_t = Cov(p_t).$$

So $A_0 = \mathrm{Id}$, and as a side remark we know that $A_t \leq \mathrm{Id}/t$ since we are more log-concave than the Gaussian. The main question is whether A_t can explode in a short time interval.

Proposition ("Basic estimate")**.** *For any fixed* $0 < T < 1$*, if*

$$\mathbb{E} \int_0^T \|A_t\|_{op} dt \leq \frac{1}{8}$$

then $\psi_{p_0} \geq c\sqrt{T}$*.*

Proof of basic estimate. Fix $E \subset \mathbb{R}^n$ with $p(E) = 1/2$, say with smooth boundary. Then by the martingale property,

$$p^+(\partial E) = \mathbb{E} p_t^+(\partial E) \geq c\sqrt{t} \cdot \mathbb{E} p_t(E) \cdot (1 - p_t(E)) = c\sqrt{t}\mathbb{E} M_t(1 - M_t),$$

where $M_t = p_t(E) = \int_E p_t(x)dx$. The process $(M_t)_{t \geq 0}$ is a martingale:

$$dM_t = \int_E \langle x - b_t, dW_t \rangle p_t(x) dx = \langle v_t, dW_t \rangle$$

with $v_t = \int_E (x - b_t) d\mu_t$. Now,

$$|v_t| = \sup_{\theta \in S^{n-1}} \int_E \langle x - b_t, \theta \rangle p_t(x) dx \le \sup_{\theta \in S^{n-1}} \sqrt{\int_{\mathbb{R}^n} \langle x - b_t, \theta \rangle^2 p_t(x) dx} = \sqrt{\|A_t\|_{op}}.$$

Apply the Itô formula

$$d\left[M_t(1 - M_t)\right] = -|v_t|^2 dt + \text{martingale term} = -\|A_t\|_{op} + \text{martingale term},$$

where “martingale term” means dN_t for some martingale $(N_t)_{t \ge 0}$. Therefore,

$$\mathbb{E} M_T(1 - M_T) = M_0(1 - M_0) - \mathbb{E} \int_0^T \|A_t\|_{op} dt \ge \frac{1}{4} - \frac{1}{8} = \frac{1}{8}.$$

□

Remark: In Lecture 3 we will show that the basic estimate is tight, up to logarithmic factors. That is, up to a log-factor, the squared KLS constant ψ_n^2 is the maximal time where all eigenvalues of the covariance matrix are bounded by a universal constant.

Example: Suppose that $p = p_0$ is a product measure. In this case the eigenvalues $\lambda_1, \ldots, \lambda_n$ are independent stochastic processes. If you compute their tail distribution, it turns out that one of them can reach $\log n$ in a short time, a bit like exponential random variables, and the operator norm would be at least $\log n$. In fact, in the example when p is a product of exponential random variables, in time $t = C / \log n$ for appropriate constant $C > 0$ we have $\mathbb{E}\|A_t\|_{op} \sim \log n$.

Dynamics of the covariance matrix. First, what is the dynamics of the barycenter?

$$db_t = \int_{\mathbb{R}^n} x \langle x - b_t, dW_t \rangle p_t(x) dx = \left[\int_{\mathbb{R}^n} x \otimes (x - b_t) p_t(x)\right] dW_t = A_t dW_t.$$

Next we claim that

$$dA_t = \left[\int_{\mathbb{R}^n} (x - b_t)^{\otimes 3} p_t(x) dx\right] dW_t - A_t^2 dt. \tag{3}$$

That is, a symmetric 3-tensor is contracted with a vector and this gives a matrix, something like $\sum a_{ijk} v_k$. Indeed, to prove (3), by translation invariance it suffices to consider the case where $b_t = 0$. In this case the computation is as in the case of the barycenter, yet there is also an Itô term coming from the stochastic differential $b_t \otimes b_t$, which is $-A_t^2 dt$. Equation (3) means that the (i, j)-entry of the matrix A_t satisfies

$$d(A_t)_{i,j} = \langle \xi_{i,j}, dW_t \rangle - (A_t^2)_{i,j} dt$$

where

$$\xi_{i,j} = \int_{\mathbb{R}^n} x x_i x_j p_t(x + b_t) dx.$$

Lecture 3 – Moments of order 3 along the dynamics

Here we incorporate simplifications suggested in discussions with **Daniel Dadush**, **Ronen Eldan** and **Joseph Lehec**. We study the moments of the covariance matrix A_t. For a positive integer $q \geq 3$ set

$$\Gamma_t = \mathrm{Tr}[A_t^q].$$

We need to show that this does not grow too quickly. Two types of controls are useful:

(A) Bound in terms of t:

$$d\Gamma_t \leq 2q^2 \cdot \frac{1}{t}\Gamma_t dt + \text{martingale term}.$$

(B) In terms of ψ_n:

$$d\Gamma_t \leq Cq^2 \cdot \psi_n^{-2} \|A_t\|_{op} \cdot \Gamma_t + \text{martingale term}.$$

(It's not needed here, but the martingale terms are $v_t \cdot dW_t$ with $|v_t| \leq q\Gamma_t \sqrt{\|A_t\|_{op}}$).

Write $\|A_t\|_q = \Gamma_t^{1/q} = Tr[A_t^q]^{1/q}$ for the q-Schatten norms.

Corollary of bound A (Power-law growth – a new ingredient by Chen). For $t > s$,

$$\mathbb{E}\|A_t\|_q \leq \left(\frac{t}{s}\right)^{2q} \mathbb{E}\|A_s\|_q.$$

Proof. Recall that $d\Gamma_t \leq (2q^2/t)\Gamma_t dt + \text{martingale term}$. The function $x^{1/q}$ is concave, hence it yields negative Itô terms and

$$d\Gamma_t^{1/q} \leq \frac{1}{q}\Gamma_t^{1/q-1} d\Gamma_t \leq \frac{1}{q}\Gamma_t^{1/q-1} \cdot (2q^2/t)\Gamma_t dt + \text{martingale term}.$$

Therefore

$$\frac{d\mathbb{E}\|A_t\|_q}{dt} \leq \frac{2q}{t}$$

and

$$\log \frac{\mathbb{E}\|A_t\|_q}{\mathbb{E}\|A_s\|_q} \leq \log \left(\frac{t}{s}\right)^{2q}.$$

□

Corollary of bound B ("The basic estimate from Lecture 2 is sharp, up to log factors"). There exists a small enough $c > 0$ such that for any $T > 0$ (deterministic),

$$T \leq c\frac{\psi_n^2}{\log n} \qquad \Longrightarrow \qquad \mathbb{E}\|A_T\|_{op} \leq 3$$

and moreover $\mathbb{E}\|A_T\|_q \leq 3n^{1/q}$ for all $q \geq 1$.

Proof. Enough to show that $\mathbb{P}\left(\|A_T\|_{op} < 2\right) \geq 1 - 1/n^{10}$, since p is supported in $B(0, n^5)$ and $\|A_T\|_{op} \leq n^{10}$. Let

$$\tau = \inf\{t > 0\,;\, \|A_t\|_{op} \geq 2\},$$

which is a *stopping time*. By the optional stopping theorem, if M_t is a martingale, then also $M_{t\wedge\tau}$ is a martingale. Set $q = \lceil 40 \log n \rceil$ and denote

$$X_t = \Gamma_{t\wedge\tau}.$$

Then we have

$$dX_t \leq C \log^2 n \cdot \psi_n^{-2} \|A_t\|_{op} \cdot X_t dt + dM_t \leq \tilde{C} \log^2 n \cdot \psi_n^{-2} \cdot X_t dt + dM_t$$

for a martingale M_t. Therefore,

$$\frac{d}{dt}\mathbb{E}X_t \leq \tilde{C} \log^2 n \cdot \psi_n^{-2} \cdot \mathbb{E}X_t.$$

Since $X_0 = \Gamma_0 = n$,

$$\mathbb{E}X_T \leq n \cdot e^{\tilde{C}\log^2 n \cdot \psi_n^{-2} \cdot T} \leq n \cdot e^{\log n} = n^2$$

by the choice of T. Now,

$$n^2 \geq \mathbb{E}X_T \geq \mathbb{P}\left(\|A_T\|_{op} > 2\right) \cdot 2^{40 \log n}$$

and we obtain the right bound. □

<u>Proof of Chen's theorem</u>

It will be bootstrapping for ψ_n. Take $T_0 = c\psi_n^2 / \log n \leq 1/100$, so that $\mathbb{E}\|A_t\|_{op} \leq 2$ for $t \leq T_0$. Hence, for $t \geq T_0$ and $q \geq 2$,

$$\mathbb{E}\|A_t\|_q \leq \left(\frac{t}{T_0}\right)^{2q} \mathbb{E}\|A_{T_0}\|_q \leq 3n^{1/q} \cdot \left(\frac{t}{T_0}\right)^{2q}.$$

Therefore, for any fixed $T_1 > T_0$,

$$\begin{aligned} \int_0^{T_1} \mathbb{E}\|A_t\|_{op} dt &\leq 3T_0 + \int_{T_0}^{T_1} \mathbb{E}\|A_t\|_q dt \\ &\leq \frac{3}{100} + 3n^{1/q} \frac{T_1^{2q+1}}{T_0^{2q}}. \end{aligned}$$

Recall the basic estimate from Lecture 2: If the integral up to T_1 is at most $1/8$ for any initial data p_0 which is log-concave and isotropic, then $\psi_n \gtrsim \sqrt{T_1}$.

Can we take $T_1 \gg T_0$, and still bound the integral by $1/8$? We need to solve

$$n^{1/q}\frac{T_1^{2q+1}}{T_0^{2q}} = \frac{1}{50}.$$

This means that we should take

$$T_1 \sim T_0^{\frac{2q}{2q+1}} n^{-\frac{1}{q(2q+1)}}.$$

This yields

$$\psi_n \gtrsim \sqrt{T_1} \sim T_0^{\frac{q}{2q+1}} n^{-\frac{1}{2q(2q+1)}}.$$

Substituting the value for T_0,

$$\psi_n \gtrsim \psi_n^{\frac{2q}{2q+1}} (\log n)^{-\frac{q}{2q+1}} n^{-\frac{1}{2q(2q+1)}}.$$

Thus we bounded ψ_n by a different power of ψ_n, this is the "bootstrapping". Consequently, for any $q \geq 2$,

$$\psi_n \geq \left(\frac{c}{\log n}\right)^q n^{-1/(2q)}.$$

Taking $q \sim \sqrt{\log n / \log\log n}$ we obtain

$$\psi_n \gtrsim c_1 e^{-c_2\sqrt{\log n \log\log n}}.$$

What remains to prove are the bounds A and B for the growth of the q-norm of the covariance matrix, which involve 3-tensors. Recall the formula

$$d(A_t)_{i,j} = \langle \xi_{i,j}, dW_t\rangle - (A_t^2)_{i,j}dt$$

where

$$\xi_{i,j} = \int_{\mathbb{R}^n} x x_i x_j p_t(x+b_t)dx.$$

It is a bit easier to work in the orthonormal basis of eigenvectors of A_t. Itô's formula yields:

Computation. Write $0 < \lambda_1 \leq \lambda_2 \leq \dots \lambda_n$ for the eigenvalues of A_t, and $e_1, \dots, e_n$ are orthonormal eigenvectors. Then, for any smooth function f,

$$d\sum_{i=1}^n f(\lambda_i) = \left[\frac{1}{2}\cdot\sum_{i,j=1}^n |\xi_{i,j}|^2 \frac{f'(\lambda_i) - f'(\lambda_j)}{\lambda_i - \lambda_j} - \sum_{i=1}^n \lambda_i^2 f'(\lambda_i)\right] dt \;+\; \text{martingale term.}$$

(when the denominator $\lambda_i - \lambda_j$ vanishes, the quotient is interpreted by continuity as $f''(\lambda_i)$). The $\xi_{i,j}$ are expressed in the basis of the eigenvectors.

This computation boils down to the following linear algebra lemma.

Lemma (Similar to Hadamard variational formula, Daleckii-Krein '51) For symmetric matrices $A, H \in \mathbb{R}^{n\times n}$ and for small ε,

$$\text{Tr}f(A+\varepsilon H) = \text{Tr}f(A) + \varepsilon\cdot\text{Tr}[f'(A)H] + \frac{\varepsilon^2}{2}\cdot\text{Tr}[(g^{(1)}(A)\circ H)H] + o(\varepsilon^2)$$

where $\circ$ is the Schur product or Hadamard product, and $g^{(1)}(A)$ is the matrix whose entries in the basis of the eigenvectors of A are $(g(\lambda_i)-g(\lambda_j))/(\lambda_i-\lambda_j)$ for $g = f'$.

Proof of bounds (A) and (B). Denote $d\Gamma_t = \delta_t dt +$ martingale term. We need to prove that

$$\delta_t \le 2q^2\cdot\max\left\{\frac{1}{t}, C\psi_n^{-2}\|A_t\|_{op}\right\}\cdot\Gamma_t$$

almost surely. From now on we prove a deterministic statement, there will be no more expectations with respect to the Brownian motion. Denote the above maximum by $1/\kappa$.

Observation. The Poincaré constant (spectral gap) of p_t is at least κ. Indeed, p_t is more log-concave than $\exp(-t|x|^2/2)$, hence its Poincaré constant is at least t. By the definition of ψ_n and the Cheeger inequality, the Poincaré constant is at least $c\psi_n^2/\|A_t\|_{op}$.

Since $\Gamma_t = \text{Tr}(A_t)^q$ we need to plug in $f(t) = t^q$ in the above computation in order to find δ_t, the drift term. Substituting $f(t) = t^q$ in the computation yields

$$\delta_t \le \frac{q(q-1)}{2}\sum_{i,j=1}^n \lambda_i^{q-2}|\xi_{i,j}|^2,$$

since $(\lambda_i^{q-1}-\lambda_j^{q-1})/(\lambda_i-\lambda_j) \le (q-1)\cdot(\lambda_i^{q-2}+\lambda_j^{q-2})/2$. Now set

$$\xi_{ijk} = \mathbb{E}X_iX_jX_k$$

where X is distributed like $p_t(\cdot + b_t)$, so it is centered. So we have

$$\xi_{ij} = (\xi_{ijk})_{k=1,\dots,n}\in\mathbb{R}^n \qquad \xi_i = (\xi_{ijk})_{j,k=1,\dots,n} = \mathbb{E}X_iX\otimes X\in\mathbb{R}^{n\times n}.$$

Apply the Cauchy-Schwartz inequality and use $\mathbb{E}X_i = 0$ to obtain

$$\text{Tr}[\xi_i^2] = \text{Tr}[\xi_i\mathbb{E}X_iX\otimes X] = \mathbb{E}X_i\langle\xi_iX, X\rangle \le \sqrt{\mathbb{E}X_i^2}\cdot\sqrt{Var(\langle\xi_iX,X\rangle)}.$$

By the definition of the Poincaré constant, for $\varphi(x) = \langle\xi_iX, X\rangle$

$$Var\varphi(X) \le \frac{1}{\kappa}\cdot\mathbb{E}|\nabla\varphi(X)|^2 = \frac{4}{\kappa}\cdot\mathbb{E}|\xi_iX|^2 = \frac{4}{\kappa}\cdot\text{Tr}[A_t\xi_i^2].$$

To summarize, that Cauchy-Schwartz and the Poincaré inequality tell us that

$$\text{Tr}[\xi_i^2] \le \frac{2}{\sqrt{\kappa}}\cdot\sqrt{\lambda_i}\sqrt{\sum_{j,k=1}^n\lambda_j\xi_{ijk}^2}$$

Consequently,

$$\sum_{i=1}^{n} \lambda_i^{q-2} \mathrm{Tr}[\xi_i^2] \leq \frac{2}{\sqrt{\kappa}} \cdot \sum_{i=1}^{n} \lambda_i^{q-3/2} \sqrt{\sum_{j,k=1}^{n} \lambda_j \xi_{ijk}^2}.$$

Now apply a Cauchy-Schwartz inequality,

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i^{q-2} \mathrm{Tr}[\xi_i^2] &\leq \frac{2}{\sqrt{\kappa}} \cdot \sqrt{\sum_{i=1}^{n} \lambda_i^q} \sqrt{\sum_{i,j,k} \lambda_i^{q-3} \lambda_j \xi_{ijk}^2} \\ &\leq \frac{2}{\sqrt{\kappa}} \cdot \sqrt{\sum_{i=1}^{n} \lambda_i^q} \sqrt{\sum_{i,j,k} \lambda_i^{q-2} \xi_{ijk}^2} \\ &= \frac{2}{\sqrt{\kappa}} \cdot \sqrt{\sum_{i=1}^{n} \lambda_i^q} \sqrt{\sum_{i=1}^{n} \lambda_i^{q-2} \mathrm{Tr}[\xi_i^2]}. \end{aligned}$$

where we used that $x^{q-3}y + y^{q-3}x \leq x^{q-2} + y^{q-2}$. Therefore,

$$\sum_{i=1}^{n} \lambda_i^{q-2} \mathrm{Tr}[\xi_i^2] \leq \frac{4}{\kappa} \sum_{i=1}^{n} \lambda_i^q$$

Consequently,

$$\delta_t \leq \frac{q(q-1)}{2} \sum_{i=1}^{n} \lambda_i^{q-2} \mathrm{Tr}[\xi_i^2] \leq \frac{2q(q-1)}{\kappa} \Gamma_t,$$

as promised, even with $q(q-1)$ in place of q^2.

This completes the proof.