Lecture 3. The covariance of the localization process

Our main task for today's lecture is to prove the following.

Theorem 1. *Let μ be log-concave and isotropic, in the sense that it is centered with identity covariance matrix. Then the corresponding stochastic localization process satisfies*

$$\mathbb{E}\|\mathrm{cov}(\mu_t)\| \leq C, \quad \forall t \leq \frac{1}{C \log^2 n},$$

As usual C is a universal constant ($C = 10$ is probably OK). Also $\mathrm{cov}(\mu_t)$ is the covariance matrix of the measure μ_t. Recall that μ_t is a random measure, so covariance is a random matrix. Lastly the norm here is the operator norm, which is also the maximal eigenvalue (a covariance matrix is psd).

This can be derived by combining arguments from Eldan [2], Lee-Vempala [7], Chen [1], Klartag-Lehec [5], with the improved Lichnerowicz of Klartag [4]. Actually the improved Lichnerowicz allows to bypass many ideas of the aforementioned papers.

We have seen that μ_t is just the conditional of law of a random vector X having law μ given some noisy version of X. So we could just as well rewrite the theorem as follows. Let X be isotropic and log-concave, let G be a standard Gaussian vector independent of X, then

$$\mathbb{E}\|\mathrm{cov}(X \mid X + \sqrt{s} \cdot G)\| \leq C, \quad \forall \rho \geq C \log^2 n.$$

This morning Bo'az showed us, using something called the improved Lichnerowicz inequality, that if X is log-concave and if

$$\mathbb{E}\|\mathrm{cov}(X \mid X + \sqrt{s} \cdot G)\| \lesssim 1,$$

for all $s \geq s_0$ then $C_P(X) \lesssim \sqrt{s_0}$. So our theorem indeed yields

$$C_P(X) \lesssim \log n,$$

which is the current best bound for the Poincaré constant of an isotropic log-concave random vector.

The point of reversing time and of writing everything in terms of the stochastic localization is that we can then control the evolution of $\mathrm{cov}(\mu_t)$ using Itô's formula and some convexity inequalities. The proof of the theorem requires some preliminaries. There will be a number of them, but I promise you that taken individually, each one of them is pretty easy.

The equation for the covariance. As we have seen in the previous lecture, for any test function f the martingale $M_t = \int_{\mathbb{R}^n} f \, d\mu_t$ satisfies

$$dM_t = \left(\int_{\mathbb{R}^n} f(x)(x - a_t) \, d\mu_t \right) \cdot dW_t,$$

where (W_t) is some standard Brownian motion. This obviously extends to vector valued functions. If $F \colon \mathbb{R}^n \to \mathbb{R}^k$ is a vector valued function that grows fairly reasonably at infinity then the process (M_t) given by

$$M_t = \int F \, d\mu_t$$

is a martingale, and moreover

$$dM_t = \left(\int F(x) \otimes (x - a_t)\, d\mu_t \right) \cdot dW_t$$

A bit more explicitly, writing x_i for the i-th coordinate of a vector $x \in \mathbb{R}^n$ we have

$$dM_t = \sum_{i=1}^{n} \left(\int F(x)(x - a_t)_i\, d\mu_t \right) dW_{t,i}. \tag{1}$$

Lemma 2. *Let a_t and A_t be the barcenter and covariance matrix of μ_t, respectively. Then*

$$\begin{aligned} da_t &= A_t dW_t \\ dA_t &= \sum_{i=1}^{n} \left(\int (x - a_t)^{\otimes 2} (x - a_t)_i\, d\mu_t \right) dW_{t,i} - A_t^2\, dt. \end{aligned}$$

This is obtained by applying (1) to the tensors $F(x) = x$ and $F(x) = x \otimes x$ and then rearranging the terms appropriately. The details are left as an exercise.
This shows that the stochastic localization process has some moment generating property. The derivative for the barycenter is expressed in terms of the covariance, and the derivative for the covariance depends on 3-tensors.

Some matrix inequalities.

Lemma 3. *Suppose K, H are symmetric matrices, and K is positive semi-definite. Then for every positive α, β we have*

$$\operatorname{tr}(K^\alpha H K^\beta H) \le \operatorname{tr}(K^{\alpha+\beta} H^2).$$

Proof. Let $K = \sum \lambda_i x_i \otimes x_i$ be the spectral decomposition of K. Then

$$\begin{aligned} \operatorname{tr}(K^\alpha H K^\beta H) &= \sum_{ij} \lambda_i^\alpha \lambda_j^\beta \langle x_i, H x_j \rangle^2 \\ &\le \sum_{ij} \lambda_i^{\alpha+\beta} \langle x_i, H x_j \rangle^2 \\ &= \sum_{i} \lambda_i^{\alpha+\beta} \langle x_i, H x_i \rangle^2 = \operatorname{tr}(K^{\alpha+\beta} H^2). \end{aligned}$$

The inequality above follows from Young's inequality

$$\lambda_i^\alpha \lambda_j^\beta \le \frac{\alpha}{\alpha+\beta} \lambda_i^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} \lambda_j^{\alpha+\beta},$$

and the fact that the expression $\langle x_i, H x_j \rangle^2$ is symmetric in i and j. □

Corollary 4. *Let φ be the map defined on the space $S_n(\mathbb{R})$ of symmetric matrices by $\varphi(A) = \operatorname{tr} \mathrm{e}^A$. Then for every symmetric matrices A, H we have*

$$D^2\varphi(A)(H, H) \le D\varphi(A)(H^2) = \operatorname{tr}(\mathrm{e}^A H^2),$$

where $D\varphi(A)$ stands for the differential of φ at A and $D^2\varphi(A)$ for the Hessian matrix, viewed as a bilinear form on $S_n(\mathbb{R})$.

Proof. Assume first that the matrix A is positive. Then by the previous lemma we have

$$\begin{aligned} D^2\varphi(A)(H,H) &= \sum_{k\geq 1} \frac{1}{k!} \sum_{l=0}^{k-1} \operatorname{tr}(A^l H A^{k-1-l} H) \\ &\leq \sum_{k\geq 1} \frac{1}{k!} \cdot k \cdot \operatorname{tr}(A^{k-1}H^2) = \operatorname{tr}(\mathrm{e}^A H^2) = D\varphi(A)(H^2), \end{aligned}$$

which is the desired inequality. This argument does not work if A has some negative eigenvalues, but observe that the function φ has the property that

$$\varphi(A + t\cdot \mathrm{Id}) = \mathrm{e}^t \varphi(A)$$

By differentiating this equality with respect to A we see also $D\varphi$ and $D^2\varphi$ satisfy the same equation, which means that adding a multiple of identity to A does not perturb the desired inequality. Therefore it is enough to prove it for positive A. □

Inequalities for 3-tensors. Recall the equation for A_t

$$dA_t = \sum_{i=1}^n H_{i,t} dW_i - A_t^2\, dt,$$

where

$$H_{i,t} = \int_{\mathbb{R}^n} (x-a_t)^{\otimes 2} (x-a_t)_i \, d\mu_t.$$

Recall that a_t is the barycenter of μ_t. So the matrix $H_{i,t}$ is of the form $\mathbb{E} X_i X^{\otimes 2}$ for some random vector with mean 0. We need to control such quantities. This is the purpose of the next two lemmas.

Lemma 5. *Let X be a centered log-concave vector, and let u be a fixed unit vector. Then*

$$\sup_{u\in\mathbb{S}^{n-1}} \{\|\mathbb{E}(X\cdot u)X^{\otimes 2}\|\} \leq C\|\operatorname{cov}(X)\|^{3/2}.$$

Proof. Let u, v be unit vector and let $H_u = \mathbb{E}(X\cdot u)X^{\otimes 2}$. By Cauchy-Scwharz

$$H_u v\cdot v = \mathbb{E}(X\cdot u)(X\cdot v)^2 \leq (\mathbb{E}(X\cdot u)^2)^{1/2} (\mathbb{E}(X\cdot v)^4)^{1/2}.$$

Now we use log-concavity. The variable $X\cdot v$ is a log-concave random variable centered at 0. As we saw in Bo'az's first talk its moments satisfy a reverse Hölder inequality. In particular the fourth moment and second moment squared are of the same order. We thus get

$$H_u v\cdot v \leq C(\mathbb{E}(X\cdot u)^2)^{1/2} \mathbb{E}(X\cdot v)^2 \leq C\|\operatorname{cov}(X)\|^{3/2}.$$

Taking the supremum in both u and v yields the result. □

Lemma 6. *Let X be a centered random vector having third moments and finite Poincaré constant. Then*

$$\|\sum_{i=1}^n (\mathbb{E} X_i X^{\otimes 2})^2\| \leq 4 C_P(X)\cdot \|\operatorname{cov}(X)\|^2.$$

Proof. Recall the definition of H_u. When u is a coordinate vector e_i we write H_i rather than H_{e_i}. We need to show that for every unit vector u

$$\sum_{i=1}^n H_i^2 u \cdot u \leq 4C_P(X) \cdot \|\mathrm{cov}(X)\|^2.$$

An elementary computation shows that $\sum H_i^2 u \cdot u = \mathrm{tr}(H_u^2)$. Moreover, since X is centered, we get from Cauchy-Schwarz and the Poincaré inequality

$$\begin{aligned}
\mathrm{tr} H_u^2 &= \mathbb{E}(X \cdot u)(H_u X \cdot X) \\
&\leq (\mathbb{E}(X \cdot u)^2)^{1/2} \cdot (\mathrm{var}(H_u X \cdot X))^{1/2} \\
&\leq (\mathbb{E}(X \cdot u)^2)^{1/2} \cdot (4C_P(X)\mathbb{E}|H_u X|^2)^{1/2} \\
&= (\mathrm{cov}(X)u \cdot u)^{1/2} \cdot (4C_P(X)\mathrm{tr}(H_u^2 \mathrm{cov}(X)))^{1/2} \\
&\leq \|\mathrm{cov}(X)\| \cdot (4C_P(X)\mathrm{tr}(H_u^2))^{1/2}.
\end{aligned}$$

Thus $\mathrm{tr} H_u^2 \leq 4C_P(X)\|\mathrm{cov}(X)\|^2$, which is the result. □

Remark. We only applied Poincaré to a quadratic form so in a sense we only need a weak notion of Poincaré here. This observation will not be needed in the subsequent analysis presented here but it was crucial in the original work of Eldan.

Freedman inequality. Lastly we need a relatively classical deviation inequality for martingales, which is usually attributed to Freedman [3].

Lemma 7. *Let $(M_t)_{t\geq 0}$ be a continuous local martingale satisfying $M_0 = 0$. Then for every positive u and σ^2 we have*

$$\mathbb{P}(\exists t > 0 \colon M_t \geq u \,\&\, \langle M \rangle_t \leq \sigma^2) \leq \mathrm{e}^{-u^2/2\sigma^2}.$$

Proof. We only sketch the argument and leave the details as an exercise. Start by proving the following statement: If (Z_t) is a square integrable martingale satisfying $\langle Z \rangle_t \leq \sigma^2$ for all $t > 0$ and almost surely, then $Z_\infty = \lim_{t\to+\infty} Z_t$ exists and satisfies

$$\mathbb{P}(Z_\infty \geq u) \leq \mathrm{e}^{-u^2/2\sigma^2}$$

for all $u > 0$. Coming back to Freedman's inequality, introduce the stopping time

$$\tau = \inf\{t > 0 \colon \langle M \rangle_t > \sigma^2\}$$

and apply the above statement to the martingale (M_t) stopped at time τ. □

The bound on the covariance matrix.

Theorem 8. *Suppose μ is log-concave and isotropic on $\mathbb{R}^n$, and let (A_t) be the covariance process of the stochastic localization associated to μ. Then*

$$\mathbb{P}\left(\exists s \leq t \colon \|A_t\| \geq 2\right) \leq \exp\left(-\frac{1}{Ct}\right), \qquad \forall t \leq \frac{1}{C \log^2 n}.$$

Remark. We will see later on that this bound is pretty much sharp.

Proof. A common method to control the norm of a symmetric random matrix A is to use the Schatten norm $(\mathrm{tr} A^p)^{1/p}$ where p is an even integer of order $\log n$ as a proxy for $\|A\|$. This is what Eldan does in his 2014 paper. For some reason we prefer to use another proxy, namely

$$h_\beta(M) := \frac{1}{\beta} \log \mathrm{tr}\, \mathrm{e}^{\beta M}.$$

Note that h_β is a smooth function. Also

$$\lambda_{\max}(M) \leq \frac{1}{\beta} \log \mathrm{tr}\, \mathrm{e}^{\beta M} \leq \lambda_{\max}(M) + \frac{\log n}{\beta}.$$

Therefore if β is of order $\log n$ then $h_\beta(M)$ is approximately the same as the maximal eigenvalue of M, up to an additive constant. Recall the equation for (A_t). From Itô's formula we get (omitting the time variable)

$$dh_\beta(A) = Dh_\beta(A) \left(\sum H_i dB_i \right) - Dh_\beta(A)(A^2)\, dt + \frac{1}{2} \sum D^2 h_\beta(A)(H_i, H_i)\, dt.$$

Let

$$M = \nabla h_\beta(A) = \frac{\mathrm{e}^{\beta A}}{\mathrm{tr}(\mathrm{e}^{\beta A})},$$

and note that this is a positive semi-definite matrix of trace 1. Using Corollary 4, we see that the second derivative of h_β satisfies

$$D^2 h_\beta(A)(H_i, H_i) \leq \beta \mathrm{tr}(M H_i^2).$$

Dropping some negative terms we finally arrive at

$$dh_\beta(A) \leq \sum \mathrm{tr}(M H_i) dB_i + \frac{\beta}{2} \mathrm{tr} \left(M \sum H_i^2 \right)\, dt.$$

Let us deal with the absolutely continuous part. Since M is positive and has trace 1, we get from Lemma 6

$$\mathrm{tr} \left(M \sum H_i^2 \right) \leq \| \sum H_i^2 \| \leq 4 C_P(\mu_t) \|A_t\|^2.$$

Recall that (μ_t) gets more and more log-concave along time. In particular if the original measure μ is log-concave then μ_t is t-uniformly log-concave, almost surely. From the improved Lichnerowicz inequality of Klartag we get

$$C_P(\mu_t) \leq \left(\frac{\|A_t\|}{t} \right)^{1/2},$$

hence

$$dh_\beta(A) \leq \sum \mathrm{tr}(M H_i) dB_i + \frac{C\beta}{\sqrt{t}} \cdot \|A_t\|^{5/2} dt.$$

Let us now bound the quadratic variation of the martingale part. For any unit vector u, letting $H_u = \sum H_i u_i$ we get from Lemma 5

$$\sum \mathrm{tr}(M H_i) u_i \leq \mathrm{tr}(M H_u) \leq \|H_u\| \leq C_0 \|A_t|^{3/2}.$$

Therefore,

$$\sum \operatorname{tr}(MH_i)^2 \le C_0^2 \|A_t\|^3.$$

Let us summarize what we have obtained so far:

$$\begin{aligned} \|A_t\| \le h_\beta(A_t) &\le h_\beta(A_0) + Z_t + \frac{\beta}{2}\int_0^t s^{-1/2}\|A_s\|^{5/2}\,ds \\ &= 1 + \frac{\log n}{\beta} + Z_t + \frac{\beta}{2}\int_0^t s^{-1/2}\|A_s\|^{5/2}\,ds \end{aligned} \tag{2}$$

where (Z_t) is a continuous martingale starting from 0 whose quadratic variation satisfies

$$\langle Z\rangle_t \le C_1 \int_0^t \|A_s\|^3\,ds. \tag{3}$$

Now choose $\beta = 2\log n$, and assume that there exists $s \le t$ such that $\|A_s\| \ge 2$. If s is the smallest such time then before time s the operator norm of A is less than 2, so by (2)

$$2 = \|A_s\| \le \frac{3}{2} + Z_s + C_2 s^{1/2}\log n \le \frac{3}{2} + Z_s + C_2 t^{1/2}\log n$$

where C_2 is some constant. If t is a sufficiently small multiple of $(\log n)^{-2}$ then the latest inequality implies that $Z_s \ge \frac{1}{4}$. Moreover, thanks to (3) we also have $\langle Z\rangle_s \le C_3 s \le C_3 t$. Therefore,

$$\mathbb{P}(\exists s \le t\colon \|A_s\| \ge 2) \le \mathbb{P}(\exists s > 0\colon Z_s \ge \frac{1}{4} \ \&\ \langle Z\rangle_s \le C_3 t).$$

We conclude with Freedmann's inequality. □

Now we prove the bound for the expectation of A_t.

Proof. Since μ_t is t-uniformly log-concave, its covariance matrix is bounded above by $(1/t)\mathrm{Id}$. This was already mentioned by Bo'az. Therefore we have $\|A_t\| \le 1/t$, almost surely. As a result

$$\mathbb{E}\|A_t\| \le 2 + \frac{1}{t}\mathbb{P}(\|A_t\| > 2).$$

Now we apply the latest theorem. Since $x \cdot \mathrm{e}^{-c_1 x}$ is a bounded function of x we indeed get $\mathbb{E}\|A_t\| = O(1)$ on the appropriate time range. □

Remark. Instead of the improved Lichnerowicz inequality, we could have bounded $C_P(\mu_t)$ by the KLS constant. Namely if C_n is the largest Poincaré constant of an isotropic log-concave measure then it is easy to see that for any log-concave X

$$C_P(X) \le C_n \|\operatorname{cov}(X)\|.$$

Therefore

$$C_P(\mu_t) \le C_n \|A_t\|.$$

Using this estimate instead of the improved Lichnerowicz inequality leads to the following statement:

$$\mathbb{E}\|A_t\| = O(1), \quad \forall t \le \frac{c}{C_n \log n}$$

It looks like we are chasing our tail here: the bound for $\|A_t\|$ depends on C_n which we wanted to bound in the first place. But recall that Bo'az showed us that if t_0 is such that $\mathbb{E}\|A_t\| = O(1)$ up until time t_0 then $C_P(\mu) = O(t_0^{-1/2})$. So the latest display actually gives

$$C_n = O(\sqrt{C_n \log n})$$

which indeed recovers $C_n = O(\log n)$.

Material not really covered

Life before improved Lichnerowicz. The improved Lichnerowicz estimate is only from 2023, and it was not available to Eldan, Lee-Vempala, Chen, Klartag-Lehec. Here I will only say a few words about the original argument of Eldan.
This morning Bo'az showed us that if μ is log-concave and such that $\mathbb{E}\|A_t\| = O(1)$ up until some small time t_0 then $C_P(\mu) = O(t_0^{-1/2})$. Let us first reprove this in a slightly different manner than what Bo'az did. Let f be the function given by E. Milman's result from my first lecture. Namely f is 1-Lipschitz and such that

$$\mathrm{var}_\mu(f) \approx \|f\|_\infty^2 \approx C_P(\mu).$$

By the decomposition of variance

$$\mathrm{var}_\mu(f) = \mathbb{E}\mathrm{var}_{\mu_t}(f) + \mathrm{var}\left(\int_{\mathbb{R}^n} f\, d\mu_t\right)$$

For the first term we proceed in the same way as Bo'az: by improved Lichnerowicz and since f is 1-Lipschitz, we have

$$\mathbb{E}\mathrm{var}_{\mu_t}(f) \leq \frac{\mathbb{E}\|A_t\|^{1/2}}{\sqrt{t}}.$$

For the second term, we proceed differently. The process $M_t = \int f\, d\mu_t$ is a martingale, whose derivative is

$$dM_t = \left(\int_{\mathbb{R}^n} f(x)(x-a_t)\, d\mu_t\right) \cdot dW_t.$$

Since $\|f\|_\infty^2 \lesssim C_P(\mu)$ we get from Cauchy-Schwarz

$$\left|\int_{\mathbb{R}^n} f(x)(x-a_t)\, d\mu_t\right|^2 \lesssim C_P(\mu)\|A_t\|.$$

Hence

$$\mathrm{var}\left(\int f\, d\mu_t\right) \lesssim C_P(\mu) \int_0^t \mathbb{E}\|A_s\|\, ds.$$

If $\mathbb{E}\|A_t\| = O(1)$ up to time t_0 we finally get

$$C_P(\mu) \lesssim t^{-1/2} + t \cdot C_P(\mu),$$

for all $t \leq t_0$, which indeed implies $C_P(\mu) = O(t_0^{-1/2})$ (assuming $t_0 = O(1)$). One thing that we can notice from this proof is that if we replace the improved Lichnerowicz inequality by the usual one, namely $C_P(\mu) \leq 1/t$ in the t-uniformly log-concave case, we also get something non trivial, namely

$$C_P(\mu) = O(t_0^{-1}). \tag{4}$$

This is obviously a lot worse than what we get from improved Lichnerowicz, but still non trivial. As a matter of fact, all the aforementioned works on the KLS conjecture (prior to the latest one by Bo'az in which the improved Lichnerowicz inequality is established) rely on this estimate, one way or another. The other argument to get Poincaré from the bound on the covariance that Bo'az showed this morning may be more elementary and more natural in a way, but it only works if one happens to know the improved Lichnerowicz inequality. If you combine it with the usual Lichnerowicz inequality you get nothing. Define the constants K_n and S_n by

$$K_n = \sup\left\{ \| \sum_{i=1}^n (\mathbb{E} X_i X^{\otimes 2})^2 \| \right\}, \quad S_n = \sup\{ \frac{1}{n} \mathrm{var} |X|^2 \}$$

where both sup are taken over all log-concave isotropic random vectors on $\mathbb{R}^n$. The constant S_n is called the thin-shell constant. The thin-shell conjecture asserts that the sequence (S_n) is bounded. This is a weak form of the KLS conjecture since we only require Poincaré for a very specific function, namely the Euclidean norm squared. It was mentioned by Bo'az in his first lecture in connection with the central limit problem for convex sets. A variant of what we have done above shows that in the isotropic log-concave case we have $\mathbb{E}\|A_t\| = O(1)$ up until times $(CK_n \log n)^{-1}$. Hence from (4)

$$C_n = O(K_n \log n).$$

Moreover, by definition of S_n, given a log-concave and isotropic vector X on $\mathbb{R}^n$, a unit vector u, and an orthogonal projection P of rank k, we have $\mathbb{E}(X \cdot u)|PX|^2 \leq \sqrt{kS_k}$. Applying this to suitable chosen projections P one can estimate the eigenvalues of $\mathbb{E}(X \cdot u)X^{\otimes 2}$ and then arrive at the bound

$$K_n \lesssim \sum_{k=1}^n \frac{S_k}{k} \lesssim S_n \log n.$$

Altogether this is gives

$$C_n = O(S_n \log^2 n).$$

In other words thin-shell implies KLS up to polylog. This was the original result of Eldan.

References

[1] Chen, Y., *An Almost Constant Lower Bound of the Isoperimetric Coefficient in the KLS Conjecture.* Geom. Funct. Anal. (GAFA), Vol. 31, (2021), 34—61.

[2] Eldan, R., *Thin shell implies spectral gap via a stochastic localization scheme.* Geom. Funct. Anal. (GAFA), Vol. 23, (2013), 532–569.

[3] D.A. Freedman, *On tail probabilities for martingales*, Ann. Probab. 3 (1) (1975), 100-118.

[4] Klartag, B., *Logarithmic bounds for isoperimetry and slices of convex sets.* Preprint, arXiv:2303.14938, 2023.

[5] Klartag, B.; Lehec, J., *Bourgain's slicing problem and KLS isoperimetry up to polylog.* Geom. Funct. Anal. 32, No. 5, 1134-1159 (2022).

[6] Klartag, B., Putterman, E., *Spectral monotonicity under Gaussian convolution.* To appear in Ann. Fac. Sci. Toulouse Math.

[7] Lee Y.T.; Vempala S.S. *Eldan's Stochastic Localization and the KLS Conjecture: Isoperimetry, Concentration and Mixing.* Preprint, 2017.