Lecture notes - From stochastic calculus to geometric inequalities

Ronen Eldan

Many thanks to Alon Nishry and Bo'az Slomka for actually reading these notes, and for their many suggestions and corrections.

1 A (very informal) crash course in Itô calculus

The aim of this section is to review a few central concepts in Itô calculus. It is intended give readers who are not familiar with this subject (hence, analysts or geometers who lack the necessary background in probability) the intuition needed in order to be able to follow these notes. Allowing ourselves to be non-rigorous, the proofs in this section are heuristic, mainly in the sense that we allow ourselves to freely change between discrete and continuous time.

1.1 A discrete axis of time

In the following, we fix a small number, $dt > 0$, which denotes a time increment. Some of our formulas will only be correct in the limit $dt \to 0$, but for simplicity and for the sake of intuition, the definitions as well as some of the calculations will be carried out with dt being a fixed, small number.

Consider the lattice $\Lambda = \mathbb{Z}dt$. In the following, a stochastic process X_t parametrized by time will be defined for $t \in \Lambda$. We will understand the differential of X_t as

$$dX_t := X_{t+dt} - X_t, \quad \forall t \in \Lambda. \tag{1}$$

Moreover, by abuse of notation, for stochastic processes X_t, Y_t, Z_t and for $t_1, t_2 \in \Lambda$, we will write

$$\int_{t_1}^{t_2} (Y_t dX_t + Z_t dt) := \sum_{t \in \Lambda \cap [t_1, t_2)} \left(Y_t (X_{t+dt} - X_t) + Z_t \cdot dt \right).$$

1.2 Brownian motion and filtration

A Brownian motion W_t on $\mathbb{R}$ will then be understood as follows: We fix a doubly-infinite sequence of independent standard Gaussian random variables, $\dots, N_{-1}, N_0, N_1, \dots$ and define for every time $t \in \Lambda$,

$$dW_t = \sqrt{dt} N_{i(t)}$$

with $i(t) = t/dt$. We now define the process W_t via equation (1), setting $W_0 = 0$. With a slight abuse of notation, we'll allow ourselves to write, instead of the above,

$$W_t = \int_0^t dW_t.$$

We will also need to associate a filtration with our Brownian motion. To that end, we set $\mathcal{F}_t = \sigma(\{dW(s); s \in \Lambda \cap (-\infty, t)\})$. We will say that a process Z_t is adapted to $\mathcal{F}_t$ if, for every $t \in \Lambda$ one has that Z_t is $\mathcal{F}_t$-measurable.

1.3 Itô processes

Let $\{X_t\}_{t\in\Lambda}$ be a sequence of random variables. We say that X_t is an Itô process if there exist two processes $\{\sigma_t\}_{t\in\Lambda}$ and $\{\mu_t\}_{t\in\Lambda}$, adapted to $\mathcal{F}_t$, such that

$$X_{t+dt} - X_t = \sigma_t(W_{t+dt} - W_t) + \mu_t dt$$

or, in other words,

$$dX_t = \sigma_t dW_t + \mu_t dt, \quad \forall t \in \Lambda. \tag{2}$$

We think of σ_t as the local variance of X_t and of μ_t as the local drift. In the special case that $\mu_t = 0$, we will say that X_t is a **martingale**.

A useful fact is the following: if X_t is a martingale and Y_t is a process adapted to $\mathcal{F}_t$, we have that

$$\mathbb{E}\int_0^t Y_t dX_t = \int_0^t \mathbb{E}[Y_t dX_t] = \int_0^t \mathbb{E}\big[\mathbb{E}[Y_t dX_t|\mathcal{F}_t]\big] = \int_0^t \mathbb{E}\big[Y_t \mathbb{E}[dX_t|\mathcal{F}_t]\big] = 0. \tag{3}$$

1.4 Quadratic variation

The quadratic variation of a process X_t is defined as

$$[X]_t = \sup_{k\in\mathbb{N}} \sup_{1\le t_1 \le \cdots \le t_k \le t} \sum_{i=1}^k (X_{t_i} - X_{t_{i-1}})^2.$$

For a process X_t satisfying equation (2), we have by the law of large numbers (informally, as $dt \to 0$)

$$\begin{aligned}
[X]_t &= \sum_{s\in\Lambda\cap[0,t)} (dX_s)^2 \\
&= \sum_{s\in\Lambda\cap[0,t)} N_{i(t)}^2 \sigma_t^2 dt + \mu_t (dt)^2 \\
&= \int_0^t \sigma_t^2 dt + O(\sqrt{dt}).
\end{aligned}$$

The last equality requires some continuity assumption about the process σ_t. In the following we will allow ourselves to assume that the process σ_t is nice enough, and the increment dt is negligible so that

$$[X]_t = \int_0^t \sigma_t^2 dt.$$

1.5 Itô's formula

Let X_t be an Itô process and let $Y_t = f(X_t)$, where f is a nice function (have a continuous second derivative, say). Our goal is to find an expression for the differential dY_t, in the form of equation (2). Naively, one *might* think that by the chain rule, we simply have

$$dY_t = Y_{t+dt} - Y_t = f(X_{t+dt}) - f(X_t) = f'(X_t)dX_t + o(dt).$$

However, this is **not correct**. The reason is that, since the term dX_t is of the order $\sqrt{dt}$ rather than the order dt, there is another term we have to take into account. In other words, our chain rule has to involve the second derivative of f as well. We actually have,

$$\begin{aligned} dY_t &= Y_{t+dt} - Y_t \\ &= f(X_{t+dt}) - f(X_t) \\ &= f(X_t + \sigma_t dW_t + \mu_t dt) - f(X_t) \\ &= f'(X_t)(\sigma_t dW_t + \mu_t dt) + \frac{1}{2} f''(X_t)\sigma_t^2 (dW_t)^2 + o(dt). \end{aligned}$$

Since dW_t is of the order $\sqrt{dt}$ rather than of the order dt, we have that the second term is *not* negligible compared to dt, it is rather of the same order. This term is called the Itô term. Moreover, note that we have $\mathbb{E}[dW_t^2]/dt = \mathbb{E}N_{i(t)}^2 = 1$. The law of large numbers will allow us to (informally) write $(dW_t)^2 = dt$. We conclude that

$$dY_t = f'(X_t)\sigma_t dW_t + \left(f'(X_t)\mu_t + \frac{1}{2} f''(X_t)\sigma_t^2 \right) dt$$

which is correct, under suitable assumptions, as $dt \to 0$.

Following the same lines of proof, we can assume that the function $f = f(x,t)$ depends also on time, hence we define $Y_t = f(X_t, t)$. In this case we get an extra term:

$$dY_t = \frac{\partial}{\partial x} f(X_t, t)\sigma_t dW_t + \left(\frac{\partial}{\partial x} f_t(X_t)\mu_t + \frac{1}{2}\frac{\partial^2}{\partial x^2} f(X_t, t)\sigma_t^2 + \frac{\partial}{\partial t} f(X_t, t) \right) dt.$$

An easy, but important, consequence of Itô's formula is the following: If X_t is a martingale, by taking $f(x) = (x - \mathbb{E}[X_0])^2$ we deduce from Itô's formula that

$$df(X_t) = 2(X_t - \mathbb{E}[X_0])dX_t + \sigma_t^2 dt = 2(X_t - \mathbb{E}[X_0])dX_t + d[X]_t.$$

Since $\mathbb{E}[X_t] = \mathbb{E}[X_0]$, we now have

$$\mathrm{Var}[X_t] = \mathbb{E}[(X_t - \mathbb{E}[X_t])^2] = \mathbb{E}[f(X_t)] = [X]_t. \tag{4}$$

where we used the fact that $\mathbb{E}[dX_t] = 0$.

1.6 Higher dimensions

The generalization of the above to higher dimensions is pretty straightforward. A Brownian motion in $\mathbb{R}^n$ is a vector valued process W_t such that each coordinate is an independent one dimensional Brownian motion. For an Itô process X_t, equation (2) is changed so that σ_t is a positive semidefinite $n \times n$ matrix and μ_t takes values in $\mathbb{R}^n$ (in general, the process W_t can be in some $\mathbb{R}^m$ and σ_t can be any linear operator $\mathbb{R}^m \to \mathbb{R}^n$, but it is easy to see that one can always take σ_t to be a positive semidefinite square matrix instead).

Itô's formula becomes

$$dY_t = \langle \nabla f_t(X_t), \sigma_t dW_t \rangle + \left(\langle \nabla f_t(X_t), \mu_t \rangle + \frac{1}{2}\mathrm{Tr}\left(\sigma_t^T \mathrm{Hess} f(X_t) \sigma_t \right) \right) dt.$$

2 Two useful constructions in analysis

We would like to briefly recall two fundamental constructions: the Log-Laplace transform and the Ornstein-Uhlenbeck semigroup, along with some of their basic properties. As we will see below, the main construction discussed in these lecture notes attains several properties that resemble these two constructions.

2.1 The Log-Laplace transform

For a function $f : \mathbb{R}^n \to \mathbb{R}$, we define

$$\mathcal{L}[f](\theta) := \log \int_{\mathbb{R}^n} f(x) \exp(\langle x, \theta \rangle) dx.$$

Define also $f_\theta = \frac{f(x)\exp(\langle x,\theta\rangle)}{\int f(x)\exp(\langle x,\theta\rangle)}$, the "tilted" version of $f(x)$. The function $\theta \to \mathcal{L}[f](\theta)$ is a convex function satisfying:

- $\nabla \mathcal{L}[f](\theta) = \int x f_\theta(x) dx$ is the center of mass of f_θ.
- $\nabla^2 \mathcal{L}[f](\theta) = \mathrm{Cov}(f_\theta)$.
- In general, $\nabla^{(k)} \mathcal{L}[f](\theta)$ corresponds to the k-th cumulant tensor of f_θ.

This transform helps relate the moments of a function to its tail behaviors and entropic properties, via taking derivatives. In the past decade, it has proven extremely useful in proving bounds on the distribution of mass in convex bodies (or log-concave measures), such as **B. Klartag**'s $n^{1/4}$ bound for the isotropic constant.
See `www.newton.ac.uk/files/seminar/20110622153016302-152712.pdf` for some examples.

2.2 The Ornstein-Uhlenbeck semigroup

Denote by γ the density of the standard Gaussian measure in $\mathbb{R}^n$,

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

By slight abuse of notation, we'll also denote by γ the standard Gaussian measure. Let Γ be a random vector with law γ. Again let $f : \mathbb{R}^n \to \mathbb{R}$ be a nice enough function. For every $t > 0$, we define

$$P_t[f](x) := \mathbb{E}\left[f\left(e^{-t/2} x + \sqrt{1 - e^{-t}}\Gamma \right)\right].$$

It is characterized by the following PDE (heat flow on Gaussian space):

$$\frac{d}{dt} P_t[f](x) = (\Delta + x \cdot \nabla) P_t[f](x).$$

Let us review some of its properties. We have $\int f d\gamma = \int P_t[f] d\gamma$ for all $t \geq 0$. Thus, it makes sense to consider non-negative functions f satisfying $\int f d\gamma = 1$. Upon doing so, the Ornstein Uhlenbeck semigroup induces a flow on the space of probability measures, starting from $f d\gamma$, towards the measure γ.

Define $g_{t,x}(y) = Z_t^{-1} \exp\left(\frac{-|x-y|^2}{2(1-e^{-t})} \right)$, where Z_t is chosen so that $g_{t,x}$ is a probability measure. One has the following easy identities:

- $P_t[f](x) = \int f(y) g_{t,x}(y) dy$
- $\nabla P_t[f](x) = \int (y-x) f(y) g_{t,x}(y) dy$
- $\nabla^2 P_t[f](x) = \int (y-x)^{\otimes 2} f(y) g_{t,x}(y) dy = \mathrm{Cov}(f \cdot g_{t,x})$.
- Again, we can keep differentiating and get higher and higher moments.

3 A stochastic construction

We now define the central construction of these notes. Fix a probability measure μ on $\mathbb{R}^n$. For a function F define

$$a_\mu(F) = \int x F(x) \mu(dx) = \mathbb{E}[F\mu].$$

Let W_t be a Brownian motion in $\mathbb{R}^n$. We consider the following equation:

$$F_0(x) = 1, \;\; dF_t(x) = \langle x - a_t, dW_t \rangle F_t(x), \;\; \forall x \in \mathbb{R}^n$$

where

$$a_t = a_\mu(F_t)$$

Define also

$$A_t = \int (x - a_t)^{\otimes 2} F_t \mu(dx) = \mathrm{Cov}(F_t \mu).$$

This is an infinite system of SDE's (one equation for every $x \in \mathbb{R}^n$). We will not discuss existence and uniqueness of the solution here, but later on we'll see that by taking a slightly different point of view, this actually becomes a finite system of equations. Finally, for every $t > 0$ we define the measure μ_t by the equation

$$\frac{d\mu_t}{d\mu}(x) = F_t(x)$$

or in other words, $\mu_t = F_t \mu$.

3.1 Some basic properties of the process μ_t

Let us start with some basic things. First, we claim that μ_t is a probability measure for every $t > 0$. By allowing ourselves to freely change the order of differentiation and integration, we can write

$$d \int_{\mathbb{R}^n} \mu_t(dx) = d \int_{\mathbb{R}^n} F_t \mu(dt) = \int_{\mathbb{R}^n} dF_t(x) \mu(dx)$$

$$= \int_{\mathbb{R}^n} \langle x - a_t, dW_t \rangle \mu_t(dx) = \left\langle \int_{\mathbb{R}^n} x \mu_t(dx) - a_t, dW_t \right\rangle = 0.$$

Another straightforward fact is:

Fact 1. *For every $x \in \mathbb{R}^n$, the process $F_t(x)$ is a martingale and for every test-function φ, the process $\int \varphi d\mu_t$ is a martingale.*

Informally, our construction has a *semi-group* property, in the sense that the two following descriptions are equivalent: (a) Begin with a measure μ, run the process up to some time t, then take the measure $\nu = \mu_t$ and run the process on this measure up to time s, obtaining a measure ν_s. (b) Run the process up to time $t + s$, obtaining the measure μ_{t+s}.

3.1.1 A slightly different point of view

Another fact we need is a simple application of Itô's formula. We calculate, for every $x \in \mathbb{R}^n$

$$d \log F_t(x) = \frac{dF_t(x)}{F_t(x)} - \frac{1}{2}\frac{d[F(x)]_t}{F_t(x)^2} = \langle x - a_t, dW_t \rangle - \frac{1}{2}\left| x - a_t \right|^2 dt$$

$$= -\frac{1}{2}|x|^2 dt + \text{ some linear function in x}$$

Consequently, we have that

$$F_t(x) = Z_t^{-1} \exp\left(\langle c_t, x \rangle - \frac{1}{2} t |x|^2 \right) \tag{5}$$

for some Itô processes Z_t, c_t.

3.1.2 Convergence to a δ-measure

In view of equation (5), we have that, as $t \to \infty$, μ_t converges (weakly, with respect to continuous functions) to a Dirac δ-measure. Denoting $a_\infty = \lim_{t\to\infty} a_t$, we have for every continuous test function φ, that

$$\int \varphi(x)\mu(dx) = \lim_{t\to\infty} \mathbb{E}\left[\int \varphi(x)\mu_t(dx) \right] = \lim_{t\to\infty} \mathbb{E}[\varphi(a_t)] = \mathbb{E}[\varphi(a_\infty)].$$

In other words we have that $a_\infty \sim \mu$.

3.1.3 The process μ_t as a moment-generating process

Fix a test function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and consider the process μ_t. Define

$$M_t = \int \varphi(x)\mu_t(dx). \tag{6}$$

By Fact 1, we have that M_t is a martingale. Let us calculate the stochastic differential of M_t:

$$dM_t = d\int \varphi(x)F_t(x)\mu(dx) = \int \varphi(x) dF_t(x)\mu(dx)$$

$$= \int \varphi(x)\langle x - a_t, dW_t\rangle F_t(x)\mu(dx) = \left\langle \int \varphi(x)(x - a_t)\mu_t(dx), dW_t \right\rangle.$$

And in other words,

$$d[M]_t = \left| \int \varphi(x)(x - a_t)\mu_t(dx) \right|^2 dt. \tag{7}$$

Next, let us define

$$M_t^{(1)} = \int \varphi(x)(x - a_t)\mu_t(dx)$$

so that $dM_t = \langle M_t^{(1)}, dW_t \rangle$. A similar calculation yields

$$dM_t^{(1)} = \left(\int \varphi(x)(x - a_t)^{\otimes 2} F_t(x)\mu(dx) \right) dW_t - M_t A_t dW_t.$$

In general, defining

$$M_t^{(k)} = \int \varphi(x)(x-a_t)^{\otimes k} F_t \mu(dx)$$

we see that $dM_t^{(k)}$ involves the term $M_t^{(k+1)} dW_t$. Let us also calculate

$$da_t = d\int xF_t(x)d\mu(x) = \int x\langle x-a_t, dW_t\rangle F_t(x)d\mu(x) \tag{8}$$

$$= \int (x-a_t)\langle x-a_t, dW_t\rangle F_t(x)d\mu(x) = A_t dW_t.$$

A similar calculation (in spirit) also yields

$$dA_t = \left(\int_{\mathbb{R}^n} (x-a_t)^{\otimes 3}\mu_t(dx)\right) dW_t - A_t^2 dt. \tag{9}$$

3.2 A simple example

Let us looks at the simple case that μ is the standard Gaussian measure as an example. Namely,

$$\frac{d\mu}{dx} = (2\pi)^{-n/2} e^{-|x|^2/2}.$$

According to formula (5), the density of μ_t takes the form,

$$\frac{d\mu_t}{dx} = Z_t^{-1} \exp\left(\langle x, c_t\rangle - \frac{1}{2}(t+1)|x|^2\right)$$

where $Z_t \in R, c_t \in \mathbb{R}^n$ are Itô processes. It follows that the covariance matrix A_t satisfies

$$A_t = (t+1)^{-1} Id.$$

Next, we use (8) to derive that,

$$a_t = \int_0^t (s+1)^{-1} dW_s.$$

and finally, by (5) we know that μ_t is a Gaussian, thus we must have that

$$\frac{d\mu_t}{dx} = (t+1)^{n/2}(2\pi)^{-n/2} \exp\left(-\frac{1}{2}(t+1)\left|(x-a_t)\right|^2\right).$$

4 Towards the KLS conjecture: from moments to concentration

4.1 C-concentrated measures

We begin with the definition of a C-concentrated measure.

Definition 2. Let μ be a probability measure on $\mathbb{R}^n$. We say that μ is C-concentrated if for every 1-Lipschitz function $\varphi : \mathbb{R}^n \to \mathbb{R}$, one has

$$\sqrt{\mathrm{Var}_\mu[\varphi]} := \sqrt{\mathrm{Var}_{X\sim\mu}[\varphi(X)]} \le C.$$

We have, for example,

Fact 3. *The standard Gaussian on* $\mathbb{R}^n$ *is* 1*-concentrated.*

A remarkable result by E.Milman (relying on previous works by Buser, Cheeger, Ledoux and others) is the following.

Theorem 4. *(E. Milman, Buser, Cheeger, Ledoux etc.) (informal) Let* μ *be a log-concave measure in* $\mathbb{R}^n$*. The following are equivalent.*

(i) The measure μ *is* C*-concentrated.*

(ii) The measure μ *satisfies the following Poincaré inequality: for every smooth* φ *one has*

$$\mathrm{Var}_\mu[\varphi] \leq C^2 \int |\nabla\varphi|^2 d\mu$$

(iii) The measure μ *satisfies a Cheeger inequality of the form*

$$C\mu_+(\partial T) \geq \mu(T)(1-\mu(T)), \;\; \forall T \subset \mathbb{R}^n.$$

4.2 The KLS conjecture

The KLS conjecture roughly asserts that a log-concave measure (or alternatively, the uniform measure over a convex body) admits the same concentration as the one that the Gaussian admits. However, the fact that a measure is C-concentrated is *not* invariant to linear transformations of the measure. Thus, we need to find a suitable normalization. A natural way to do it is simply to assume that *linear* functions are concentrated:

Conjecture 5. *(Kannan-Lovász-Simonovitz) Let* $K \subset \mathbb{R}^n$ *and* μ *be uniform on* K*. Suppose that for every* linear, 1*-Lipschitz function* g*, one has* $\mathrm{Var}_\mu[g] \leq 1$*. Then* μ *is* C*-concentrated for a universal constant* $C > 0$*.*

Remark 6. Another normalization which is more common in the literature and gives an equivalent conjecture is to assume that the covariance matrix of μ is the identity.

We're interested in another conjecture, known as the *variance* conjecture or as the *thin-shell* conjecture due to Anttila, Ball and Perissinaki and Bobkov-Koldobsky. This conjecture is a special case of the KLS conjecture, and asserts that the Euclidean norm is concentrated for log-concave measures.

Conjecture 7. *(thin-shell) Let* $K \subset \mathbb{R}^n$ *and* μ *be uniform on* K*. Suppose that for every* linear, 1*-Lipschitz function* g*, one has* $\mathrm{Var}_\mu[g] \leq 1$*. Then for* $f(x) = |x|$ *one has that* $\sqrt{\mathrm{Var}_\mu[f]} \leq C$*, where* C *is a universal constant.*

The first breakthrough in attaining such a bound is due to **Klartag**, who shows that one can take $C = C_n = n^{0.49}$ in the above. The state of the art in the above bound is due to **Guédon-Milman** who show that one can take $C = C_n = n^{1/3}$.

Our theorem is, up to logarithmic factors, a reduction of the former conjecture to the latter. Denote by $\mathcal{M}_n$ the family of measures μ, which are uniform over some convex set $K \subset \mathbb{R}^n$ and satisfy the normalization condition above (i.e., for every linear 1-Lipschitz function f one has $\mathrm{Var}_\mu[f] \leq 1$). Define

$$C_n := \inf\{C; \forall\mu \in \mathcal{M}_n, \; \mu \text{ is } C\text{-concentrated}\}$$

and

$$V_n := \sup_{\mu \in \mathcal{M}_n} \sqrt{\mathrm{Var}_\mu[x \to |x|]}.$$

It is clear, by definition, that $C_n \geq V_n$. However, we also have

Theorem 8. *For all n one has $C_n \leq V_n \log n$.*

4.3 Proof of the reduction

Fix a test function $\varphi : \mathbb{R}^n \to \mathbb{R}$. Define the martingale M_t as in equation (6). Our ultimate goal is to give a bound for $\mathrm{Var}_\mu[\varphi] = \mathrm{Var}[M_\infty]$. Since M_t is a martingale, we have by orthogonality

$$\mathrm{Var}_\mu[\varphi] = \mathrm{Var}[M_t] + \mathbb{E}\mathrm{Var}[M_\infty|\mathcal{F}_t] = \mathrm{Var}[M_t] + \mathbb{E}\mathrm{Var}_{\mu_t}[\varphi].$$

Moreover, Ito's formula gives $\mathrm{Var}[M_t] = \mathbb{E}[M_t^2] - \mathbb{E}[M_t]^2 = \mathbb{E}[M]_t$, which in turn implies that

$$\mathrm{Var}_\mu[\varphi] = \mathbb{E}[M]_t + \mathbb{E}\mathrm{Var}_{\mu_t}[\varphi]. \tag{10}$$

The following well-known result (due to Brascamp-Lieb) will be useful to us:

Proposition 9. *Let $V : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $K > 0$. Suppose that,*

$$d\nu(x) = Ze^{-V(x) - \frac{1}{2K}|x|^2} dx$$

is a probability measure whose barycenter lies at the origin. Then ν satisfies a Poincaré inequality with constant K, namely for every differentiable function φ, we have

$$\mathrm{Var}_\nu[\varphi] \leq K \int |\nabla \varphi(x)|^2 \nu(dx).$$

Plugging equation (5) with the above proposition, we have

$$\mathrm{Var}_{\mu_t}[\varphi] \leq t^{-1} \int |\nabla \varphi(x)|^2 \mu_t(dx)$$

and in particular, when φ is 1-Lipschitz we have that

$$\mathrm{Var}_{\mu_t}[\varphi] \leq t^{-1}. \tag{11}$$

In order to bound the term $\mathrm{Var}[M_t]$, we use (7), according to which,

$$d[M]_t = \left| \int \varphi(x)(x - a_t)\mu_t(dx) \right|^2 dt$$

And setting $\theta = \frac{\int \varphi(x)(x-a_t)\mu_t(dx)}{\left|\int \varphi(x)(x-a_t)\mu_t(dx)\right|}$ and using Cauchy-Schwartz,

$$\begin{aligned}
d[M]_t &= \left(\int \varphi(x) \langle (x - a_t), \theta \rangle \mu_t(dx) \right)^2 \\
&= \left(\int \left(\varphi(x) - \int \varphi(x)\mu_t dx \right) \langle (x - a_t), \theta \rangle \mu_t(dx) \right)^2 \\
&\leq \left(\int \left(\varphi(x) - \int \varphi(x)\mu_t dx \right)^2 \mu_t(dx) \right) \left(\int \langle (x - a_t), \theta \rangle^2 \mu_t dx \right) \\
&= \mathrm{Var}_{\mu_t}[\varphi] \langle \theta A_t \theta \rangle \leq Y_t \|A_t\|_{OP}.
\end{aligned}$$

where we set $Y_t = \mathrm{Var}_{\mu_t}[\varphi]$. Next, we consider the stopping time

$$\tau = \sup\{t > 0;\ \|A_t\|_{OP} < 2\} \wedge 1.$$

Then by integrating the above inequality (hence using Gronwall's inequality), we deduce that

$$[M]_\tau \leq 2 \int_0^\tau Y_s ds$$

Now, according to equation (11) and since we assume that the support of μ can be assumed to be inside a ball of radius n,

$$Y_t \leq \min(t^{-1}, n^2)$$

combining the last two displays finally gives

$$[M]_\tau \leq 2 + 4 \log n.$$

Finally, equations (10) and (11) combined with the above display teach us that

$$\mathrm{Var}_\mu[\varphi] \leq 2 + 4 \log n + \mathbb{E}[\tau^{-1}]. \tag{12}$$

So, our goal becomes proving a lower bound on τ.

4.4 Bounding τ from below

Our goal is to give a probabilistic upper bound for $\|A_t\|_{OP}$ for small times t. Our starting point is equation (9), which tells us that

$$dA_t = \left(\int_{\mathbb{R}^n} (x - a_t)^{\otimes 3} \mu_t(dx) \right) dW_t - A_t^2 dt.$$

It is not hard to see that the drift term $-A_t^2 dt$ is only making our matrix smaller. Roughly speaking, we may legitimately assume that

$$dA_t = \left(\int_{\mathbb{R}^n} (x - a_t)^{\otimes 3} \mu_t(dx) \right) dW_t.$$

This shows us that the quadratic variations of the entries of A_t have to do with third moments of the measure μ_t.

Define $\lambda_1(A_t), \ldots, \lambda_n(A_t)$ to be the eigenvalues of A_t. Now, fix a time $t > 0$, and let $e_1, \ldots, e_n$ be an orthonormal basis of eigenvectors of A_t. In the following differentials, this basis will be fixed. Define also, for all t,

$$\alpha_{i,j}(t) = \langle e_i, A_t e_j \rangle$$

Denoting $\xi_{i,j} = \int_{\mathbb{R}^n} \langle x - a_t, e_i \rangle \langle x - a_t, e_j \rangle (x - a_t) \mu_t(dx)$, we have by (9) (ignoring the drift term which only helps us),

$$d\alpha_{i,j}(t) = \langle \xi_{i,j}, dW_t \rangle. \tag{13}$$

We would like to derive a formula for $d\lambda_j(t)$ at $t = t_0$. The next formulas, expressing the derivatives of eigenvalues in terms of the entries of a diagonal matrix is well known:

Lemma 10. *Let A be a diagonal matrix whose eigenvalues are distinct.*

For $i \geq j$, denote the (i,j)-th and (j,i)-th entries of A by $\alpha_{i,j}$. One has, (i)

$$\frac{\partial \lambda_i(A)}{\partial \alpha_{j,k}} = \delta_{i,j}\delta_{i,k} \tag{14}$$

(ii) Whenever $i \neq j$,

$$\frac{\partial^2 \lambda_i(A)}{\partial \alpha_{i,j}^2} = \frac{2}{\lambda_i - \lambda_j} \tag{15}$$

(iii) Whenever $(j,k) \neq (l,m)$ or $i \notin \{j,k\}$,

$$\frac{\partial^2 \lambda_i(A)}{\partial \alpha_{j,k} \partial \alpha_{l,m}} = 0$$

Combining the above lemma with formula (13), we get

$$d\lambda_i = \langle \xi_{i,i}, dW_t \rangle + \sum_{j \neq i} \frac{|\xi_{i,j}|^2}{\lambda_i - \lambda_j} dt$$

Now, consider the potential

$$S_t = \sum_{i=1}^{n} \lambda_i(t)^\alpha$$

with α and define $\tilde{\tau} = \min\{t; S_t \geq 2^\alpha\}$. It is clear that $\tau \geq \tilde{\tau}$, so it is enough to derive a lower bound for $\tilde{\tau}$.

Now, Ito's formula gives

$$\begin{aligned}
dS_t &= \alpha \sum_i \lambda_i(t)^{\alpha-1} d\lambda_i(t) + \frac{1}{2}\alpha(\alpha-1)\sum_i \lambda_i(t)^{\alpha-2} d[\lambda_i]_t \\
&= \alpha \sum_i \lambda_i(t)^{\alpha-1} \left(\langle \xi_{i,i}, dW_t \rangle + \sum_{j \neq i} \frac{|\xi_{i,j}|^2}{\lambda_i - \lambda_j} dt \right) + \frac{1}{2}\alpha(\alpha-1)\sum_i \lambda_i(t)^{\alpha-2} d[\lambda_i]_t \\
&= \alpha \sum_i \sum_{j \neq i} \lambda_i(t)^{\alpha-1} \frac{|\xi_{i,j}|^2}{\lambda_i - \lambda_j} dt + \frac{1}{2}\alpha(\alpha-1)\sum_i \lambda_i(t)^{\alpha-2} |\xi_{i,i}|^2 dt + \text{martingale} \\
&= \frac{1}{2}\alpha \sum_{i \neq j} |\xi_{i,j}|^2 \frac{\lambda_i(t)^{\alpha-1} - \lambda_j(t)^{\alpha-1}}{\lambda_i - \lambda_j} dt + \frac{1}{2}\alpha(\alpha-1)\sum_i \lambda_i(t)^{\alpha-2} |\xi_{i,i}|^2 dt + \text{martingale} \\
&\leq \frac{1}{2}\alpha(\alpha-1) \sum_{i \neq j} |\xi_{i,j}|^2 \max(\lambda_i, \lambda_j)^{\alpha-2} dt + \frac{1}{2}\alpha(\alpha-1)\sum_i \lambda_i(t)^{\alpha-2} |\xi_{i,i}|^2 dt + \text{martingale} \\
&\leq 2\alpha^2 \sum_{i,j} \lambda_i^{\alpha-2}(t) |\xi_{i,j}|^2 dt + \text{martingale.}
\end{aligned}$$

At this point, Lee-Vempala use this potential with $\alpha = 2$ to get

$$dS_t \leq \sum_{i,j} |\xi_{i,j}|^2 + \text{martingale.} \tag{16}$$

We'll later see how this rather easily achieves the $n^{1/4}$ bound.

On the contrary, we will take α to be of the order $\log n$. Recall that we'd like to find the first time for which $S_t = 2^\alpha$. For simplicity, let us ignore the martingale term and think of the above as an ODE. Let us also assume that $\|A_t\|_{OP} \geq 1$ (recall that we're trying to bound the operator norm from above, so this assumption is easy to make). For all $t \leq \tau$ we have that

$$dS_t \leq 2\alpha^2 \sum_i \lambda_i^\alpha(t) \sum_j |\xi_{i,j}|^2 dt \leq 2\alpha^2 S_t K_t dt$$

where we define

$$K_t := \sup_i \sum_j |\xi_{i,j}|^2.$$

Suppose that we could prove a deterministic bound $K_t \leq K$ for all $t \leq \tau$. Under this assumption, using Gronwall's inequality with the fact that $S_0 = n$, we would get

$$S_t \leq n \exp\left(2\alpha^2 K t\right)$$

and comparing S_t to 2^α, this would yield $t \geq \frac{\alpha \log 2 - \log n}{2\alpha^2 K}$. Taking $\alpha = 10 \log n$ then gives $t \geq \frac{1}{10K \log n}$. This is of course true only when ignoring the martingale term, but it is actually not hard to show that the same is true for the stopping time τ, namely we have

$$\mathbb{E}[\tau^{-1}] \leq 10K \log n$$

It remains to find a bound for K_t. To that end, first observe that X is a random vector such that $X + a_t$ has the law μ_t then we have

$$\sum_j |\xi_{i,j}|^2 = \sum_{j,k} \mathbb{E}[\langle X, e_i\rangle\langle X, e_j\rangle\langle X, e_k\rangle]^2 \leq \sup_{\theta \in \mathbb{S}^{n-1}} \|\mathbb{E}[\langle X, \theta\rangle X \otimes X]\|_{HS}^2$$

Now, write $Y = A_t^{-1/2} X$ and note that Y is by definition isotropic. We clearly have the bound

$$K_t \leq \|A_t\|_{OP}^3 \sup_{\theta \in \mathbb{S}^{n-1}} \|\mathbb{E}[\langle Y, \theta\rangle Y \otimes Y]\|_{HS}^2 \leq 8 \sup_{\theta \in \mathbb{S}^{n-1}} \|\mathbb{E}[\langle Y, \theta\rangle Y \otimes Y]\|_{HS}^2$$

for all $t \leq \tau$. The quantity on the right hand side is bounded in Lemma 1.6 of [E], which gives

$$K_t \leq C \sum_{k=1}^n V_k^2 / k.$$

4.4.1 The Lee-Vempala $n^{1/4}$ bound (up to a logarithmic term)

The main modification is that we now define $\tau = \inf\{t : \|A_t\|_{OP} \geq 2\sqrt{n}\} \wedge 1$. Since we now have a worse bound on $\|A_t\|_{OP}$ for $t \leq \tau$, equation (12) now becomes

$$\mathrm{Var}_\mu[\varphi] \leq n^{1/2}(\log n + 1) + \mathbb{E}[\tau^{-1}]$$

(we get this by following the exact same steps that yielded equation (12)). Now, in view of equation (16) and using Gronwall's inequality, we get $S_t \leq n + \sum_{i,j} |\xi_{i,j}|^2$ and $\|A_t\|_{OP} \leq \sqrt{S_t}$ (since we picked $\alpha = 2$). It follows that

$$\tau = \min\{t : S_t \geq 4n\} = \frac{n}{K}$$

under the assumption that K is an upper bound for the quantity $\sum_{i,j} |\xi_{i,j}|^2$. It thus remains to show that

$$\sum_{i,j} |\xi_{i,j}|^2 \leq n^{3/2}$$

which will imply that $\mathbb{E}[\tau^{-1}] \leq \sqrt{n}$. Observe that if X, Y are two independent random vectors both distributed with law μ_t, then a calculation gives that

$$\sum_{i,j} |\xi_{i,j}|^2 = \sum_{i,j,k} \left(\int_{\mathbb{R}^n} \langle x - a_t, e_i \rangle \langle x - a_t, e_j \rangle \langle x - a_t, e_k \rangle \mu_t(dx) \right)^2 = \mathbb{E} \langle X, Y \rangle^3$$

Now define $\tilde{X} = A_t^{-1/2} X$ and $\tilde{Y} = A_t^{-1/2} Y$ so that $\tilde{X}, \tilde{Y}$ are isotropic. Then

$$\mathbb{E}\langle X, Y \rangle^3 = \mathbb{E}\langle A_t \tilde{X}, \tilde{Y} \rangle^3.$$

We may clearly assume WLOG that A_t is diagonal (otherwise apply a unitary transformation). Suppose that $A_t = \mathrm{diag}(a_1, ..., a_n)$.

$$\langle A_t \tilde{X}, \tilde{Y} \rangle^3 = \left(\sum_i a_i \tilde{X}_i \tilde{Y}_i \right)^3.$$

Now, since for every two independent, isotropic log-concave random variables Z_1, Z_2 one has $Var[Z_1 Z_2] < C$, we have

$$Var[a_i \tilde{X}_i \tilde{Y}_i] \leq C a_i^2$$

so

$$\mathbb{E} \left(\sum_i a_i \tilde{X}_i \tilde{Y}_i \right)^3 \lesssim Var \left[\sum_i a_i \tilde{X}_i \tilde{Y}_i \right]^{3/2} \leq \left(C \sum_i a_i^2 \right)^{3/2} \leq C n^{3/2}.$$

where the first inequality follows from a reverse-Holder inequality for one-dimensional log-concave measures and the last follows from the definition of τ. This completes the argument.