

MEGA mini-course on LSI at IHP

DC



April 4, 2024, 10:30–12:00

Contents

1 Emergence of entropy in combinatorics, probability, statistics, analysis	1
1.1 Combinatorics	1
1.2 Probability	1
1.3 Statistics	2
1.4 Analysis	2
2 Boltzmann–Gibbs measures and free energy	2
3 Emergence of log-Sobolev : Markov diffusion processes	3
3.1 Boltzmann H-theorem, Bakry–Émery criterion, Gross hypercontractivity	3
3.2 Variational formula and tensorization	4
3.3 Log-Sobolev for Gaussian from two-points space via tensorization and CLT	5
4 Log-Sobolev and concentration of measure	6
4.1 Wigner Ensembles	7
4.2 Beta-Ensembles	8
References	8

1 Emergence of entropy in combinatorics, probability, statistics, analysis

1.1 Combinatorics

Number of microstates compatible with a macrostate, degree of freedom, disorder, volume, Stirling :

$$\frac{1}{n} \log \binom{n}{n_1, \dots, n_r} \xrightarrow[n=n_1+\dots+n_r \rightarrow \infty]{v_i = \frac{n_i}{n} \rightarrow p_i} S(p) := - \sum_{i=1}^r p_i \log(p_i).$$

If $A = \{1, \dots, r\}$ and $n = n_1 + \dots + n_r$ then $\text{Card}\{(x_1, \dots, x_n) \in A^n : \forall 1 \leq i \leq r : \sum_{k=1}^n \mathbf{1}_{x_k=i} = n_i\} = \binom{n}{n_1, \dots, n_r}$.

Boltzmann observation in kinetic gas theory at the start of statistical physics.

Shannon information theory : average length per symbol with code of optimal length.

1.2 Probability

If X_1, \dots, X_n are i.i.d. of law μ on a finite set $A = \{a_1, \dots, a_r\}$ then for all $x_1, \dots, x_n \in A$,

$$\mathbb{P}((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \prod_{i=1}^r \mu_i^{\sum_{k=1}^n \mathbf{1}_{x_k=a_i}} = \prod_{i=1}^r \mu_i^{n v_i} = e^{n \sum_{i=1}^r v_i \log \mu_i} = e^{-n(S(v) + H(v|\mu))}.$$

where we have used the Boltzmann–Shannon entropy and Kullback–Leibler divergence or relative entropy :

$$S(v) := - \sum_{i=1}^r v_i \log v_i \quad \text{and} \quad H(v|\mu) := \sum_{i=1}^r v_i \log \frac{v_i}{\mu_i} = \sum_{i=1}^r \frac{v_i}{\mu_i} \log \frac{v_i}{\mu_i} \mu_i.$$

Boltzmann–Gibbsfication. At the heart of the Sanov large deviations principle via Laplace method.

1.3 Statistics

If Y_1, \dots, Y_n are i.i.d. of law $\mu^{(\theta)}$ on a finite set A , then Fisher likelihood of data $(x_1, \dots, x_n) \in A^n$ is

$$\ell_{x_1, \dots, x_n}(\theta) := \mathbb{P}(Y_1 = x_1, \dots, Y_n = x_n) = \prod_{i=1}^n \mu_{x_i}^{(\theta)}.$$

If X_1, \dots, X_n are observed i.i.d. of law $\mu^{(\theta_*)}$ with θ_* unknown then maximum likelihood estimator is

$$\hat{\theta}_n := \arg \max_{\theta \in \Theta} \ell_{X_1, \dots, X_n}(\theta) = \arg \max_{\theta \in \Theta} \left(\frac{1}{n} \log \ell_{X_1, \dots, X_n}(\theta) \right).$$

Asymptotic analysis via law of large numbers and entropy as asymptotic contrast

$$\frac{1}{n} \log \ell_{X_1, \dots, X_n}(\theta) = \frac{1}{n} \sum_{i=1}^n \log \mu_{X_i}^{(\theta)} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \sum_{k=1}^r \mu_k^{(\theta_*)} \log \mu_k^{(\theta)} = \underbrace{-S(\mu^{(\theta_*)})}_{\text{const}} - H(\mu^{(\theta_*)} | \mu^{(\theta)}).$$

1.4 Analysis

$$f \geq 0, \quad \partial_p \|f\|_p^p = \partial_p \int f^p d\mu = \partial_p \int e^{p \log(f)} d\mu = \int f^p \log(f) d\mu = \frac{1}{p} \int f^p \log(f^p) d\mu.$$

2 Boltzmann–Gibbs measures and free energy

Boltzmann–Shannon differential or continuous entropy of a probability measure μ on \mathbb{R}^n :

$$S(\mu) := \begin{cases} - \int f(x) \log(f(x)) dx & \text{if } d\mu(x) = f(x) dx \text{ and } f \log f \in L^1(dx) \\ +\infty & \text{otherwise} \end{cases}$$

Gaussian case and Shannon exponential entropy

$$S(\mathcal{N}(m, K)) = \log \sqrt{(2\pi e)^n \det K} \quad \text{and} \quad N(\mathcal{N}(m, K)) := \frac{e^{\frac{2}{n} S(\mu)}}{2\pi e} = (\det K)^{\frac{1}{n}}.$$

Kullback–Leibler relative entropy between two probability measures on the same space :

$$H(\nu | \mu) := \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu \geq 0 \text{ with } = \text{ iff } \mu = \nu.$$

Jensen inequality (and its equality case) for strictly convex function $x \in \mathbb{R}_+ \mapsto x \log(x)$.

We take $V : \mathbb{R}^n$ or $A \rightarrow \mathbb{R}$, interpreted as an energy, such that $Z_\beta := \int e^{-\beta V(x)} dx < \infty$ for all $\beta > 0$. Maximizing $\mu \mapsto S(\mu)$ over the constraint of average energy $\int V d\mu = \nu$ gives the maximizer

$$\mu_\beta := \frac{1}{Z_\beta} e^{-\beta V} dx,$$

provided that ν is an admissible energy, typically $\nu \geq \min V$ when V has a unique global minimum. In the Gaussian case, V is quadratic while $1/\beta = \nu$ is the variance. Here dx is the Lebesgue or counting measure.

Theorem 2.1. Variational characterization : maximum entropy at fixed average energy.

$$\int V d\mu = \int V d\mu_\beta \quad \Rightarrow \quad S(\mu_\beta) - S(\mu) = H(\mu | \mu_\beta).$$

Dual point of view : instead of fixing the average energy, let us fix the inverse temperature β and introduce

$$F(\mu) := \int V d\mu - \frac{1}{\beta} S(\mu)$$

which is the Helmholtz free energy. Lagrangian point of view, the constraint is added to the functional.

$$F(\mu_\beta) = -\frac{1}{\beta} \log(Z_\beta) \quad \text{since} \quad S(\mu_\beta) = \beta \int V d\mu_\beta + \log Z_\beta.$$

Theorem 2.2. Variational characterization : minimum free energy at fixed temperature.

$$F(\mu) - F(\mu_\beta) = \frac{1}{\beta} H(\mu | \mu_\beta).$$

This explains why H is often called free energy instead of relative entropy or Kulback–Leibler divergence.

3 Emergence of log-Sobolev : Markov diffusion processes

$(X_t)_{t \geq 0}$ reversible ergodic Markov process with state space \mathbb{R}^n .

Invariant probability measure μ . Semigroup $(P_t)_{t \geq 0}$, $P_t(f)(x) = \mathbb{E}(f(X_t) | X_0 = x)$ and infinitesimal generator $L = \partial_{t=0} P_t$. We set $f_t = P_t(f)$, for $f_0 = f \geq 0$. For simplicity we focus on Langevin type diffusion process :

$$X_t = X_0 + \sqrt{2}B_t - \int_0^t \nabla V(X_s) ds, \quad L(f)(x) = \Delta f(x) - \langle \nabla V(x), \nabla f(x) \rangle, \quad d\mu(x) = \frac{1}{Z} e^{-V(x)} dx.$$

Integration by parts

$$\int f Lg d\mu = \int g Lf d\mu = - \int \langle \nabla f, \nabla g \rangle d\mu.$$

Dyson–Ornstein–Uhlenbeck example :

$$V(x) = \|x\|^2 - \beta \sum_{i < j} \log(x_i - x_j) \quad \text{with} \quad V(x) = +\infty \quad \text{outside} \quad \{x_1 \leq \dots \leq x_n\}.$$

3.1 Boltzmann H-theorem, Bakry–Émery criterion, Gross hypercontractivity

If $d\mu_0 = f d\mu$, $f \geq 0$, $\int f d\mu = 1$, then $X_t \sim \mu_t$ with $d\mu_t := f_t d\mu$. Boltzmann H-theorem by integration by parts :

$$\partial_t H(\mu_t | \mu) = \partial_t \int f_t \log(f_t) d\mu = \int (1 + \log(f_t)) d\mu = - \int \frac{|\nabla f_t|^2}{f_t} d\mu = -I(\mu_t | \mu)$$

where we have used the Fisher information

$$I(v | \mu) := \int \frac{|\nabla f|^2}{f} d\mu, \quad f := \frac{dv}{d\mu}.$$

Theorem 3.1. Exponential decay of relative entropy \Leftrightarrow LSI.

For all constant $c > 0$, the following properties are equivalent :

- (i) Exponential decay of entropy : $\forall \mu_0, \forall t \geq 0, H(\mu_t | \mu) \leq e^{-\frac{c}{4}t} H(\mu_0 | \mu)$.
- (ii) Logarithmic Sobolev Inequality (LSI) : $\forall v, H(v | \mu) \leq \frac{c}{4} I(v | \mu)$.

Proof. We get (i) from (ii) by Grönwall since $\partial_t H(\mu_t | \mu) = -I(\mu_t | \mu) \geq -\frac{c}{4} H(\mu_t | \mu)$ (we used LSI for $v = \mu_t$).

Conversely, if $\alpha(t) := e^{-\frac{c}{4}t} H(\mu_0 | \mu) - H(\mu_t | \mu)$ then $\alpha(0) = 0$, $\alpha(t) \geq 0$ for all $t \geq 0$ thus $\alpha'(0) \geq 0$ (LSI for μ_0 !). \square

Theorem 3.2. Bakry–Émery criterion for LSI : role of convexity.

If $V = \frac{1}{2\sigma^2} \|\cdot\|^2 + C$ with C convex, then μ satisfies LSI : $\forall v, H(v | \mu) \leq \frac{\sigma^2}{2} I(v | \mu)$.

In particular by taking $C \equiv 0$ we get that the Gaussian $\mathcal{N}(0, \sigma^2 I_n)$ satisfies LSI. Moreover, a multivariate Gaussian $\mathcal{N}(m, K)$ satisfies the same LSI as $\mathcal{N}(0, \|K\|_{\text{op}}^2 I_n)$ (just use the Lipschitz map $x \mapsto \sqrt{K}(x - m)$).

Proof. By using integration by parts, we get, after some algebra, a sort of H-theorem for Fisher :

$$\partial_t I(\mu_t | \mu) = -2 \int \Gamma_2(\log(f_t)) d\mu_t \quad \text{where} \quad \Gamma_2(f) := \|\nabla^2 f\|_{\text{HS}}^2 + \langle \nabla^2 V \nabla f, \nabla f \rangle \geq \frac{1}{\sigma^2} \|\nabla f\|^2.$$

By Grönwall this gives an exponential decay of Fisher information :

$$\forall \mu_0, \forall t \geq 0, \quad I(\mu_t | \mu) \leq e^{-\frac{2}{\sigma^2}t} I(\mu_0 | \mu).$$

Finally, the LSI for $\nu = \mu_0$ comes from

$$H(\mu_0 | \mu) = - \int_0^\infty \partial_t H(\mu_t | \mu) dt = \int_0^\infty I(\mu_t | \mu) dt \leq I(\mu_0 | \mu) \int_0^\infty e^{-\frac{2}{\sigma^2} t} dt = \frac{\sigma^2}{2} I(\mu_0 | \mu).$$

We also get the convexity of $t \mapsto H(\mu_t | \mu)$ via $\partial_t^2 H(\mu_t | \mu) \geq 0$, a refinement of the Boltzmann H-theorem. \square

For all $f : E \rightarrow \mathbb{R}_+$ we define (analogy with the variance, replacing x^2 with $x \log(x)$)

$$\text{Ent}_\mu(f) := \int f \log(f) d\mu - \int f d\mu \log \int f d\mu = H(\nu | \mu) \int f d\mu \quad \text{where} \quad d\nu := \frac{f}{\int f d\mu} d\mu.$$

Theorem 3.3. Leonard Gross : Hypercontractivity of Markov semigroup \Leftrightarrow LSI.

For all constant $c > 0$ the following properties are equivalent (the norms are with respect to μ) :

- (i) Hypercontractivity of semigroup : $\forall t \geq 0, \forall f, \forall p \geq 1, \|f_t\|_{p(t)} \leq \|f\|_p$ where $p(t) := 1 + (p-1)e^{\frac{4}{c}t}$
- (ii) Logarithmic Sobolev inequality (LSI) : $\forall f, \text{Ent}_\mu(f^2) \leq c \int |\nabla f|^2 d\mu.$

The term comes from the fact that $p(t) > p = p(0)$ for all $t > 0$.

Proof. Same idea as for exponential decay with this time $\alpha(t) := \log \|f_t\|_{p(t)}$.

Involves crucially the fact that $\partial_p \int f^p d\mu = \int f^p \log(f) d\mu$ for $f \geq 0$.

We can assume that $f \geq 0$ since $|f_t| \leq |f|_t$. For all $t > 0$, we find

$$\alpha'(t) = \left(\frac{1}{p(t)} \log \int f_t^{p(t)} d\mu \right)' = \frac{p'(t)}{p(t)^2} \frac{1}{\int f_t^{p(t)} d\mu} \left(\text{Ent}_\mu(f_t^{p(t)}) + \frac{p(t)^2}{p'(t)} \int (L f_t) f_t^{p(t)-1} d\mu \right).$$

Now $p(t) - 1 = \frac{c}{4} p'(t)$, while by LSI and integration by parts we get

$$\text{Ent}_\mu(g^p) \leq \frac{c}{4} \int \frac{|\nabla g^p|^2}{g^p} d\mu = -\frac{c}{4} \frac{p^2}{(p-1)} \int (Lg) g^{p-1} d\mu.$$

Therefore $\alpha'(t) \leq 0$ for any $t \geq 0$ is equivalent to LSI. \square

LSI can also be deduced geometrically from Sobolev inequality on spheres (Beckner).

LSI can also be deduced from isoperimetric inequality (Ledoux, Bobkov).

LSI is inspiring for Hamilton Ricci flow for Poincaré conjecture (Perelman).

LSI is related to transportation of measure (Talagrand, Otto–Villani, Bobkov–Gentil–Ledoux)

3.2 Variational formula and tensorization

Lemma 3.4. Variational formula.

For all probability measure μ on E and all $f : E \rightarrow \mathbb{R}_+, f \in L^1(\mu)$, we have the linearization

$$\text{Ent}_\mu(f) = \sup \left\{ \int f g d\mu : \int e^g d\mu \leq 1 \right\}, \text{ supremum achieved for } g = \log(f) - \log \int f d\mu.$$

In particular, the inequality ≤ 1 can be replaced by the equality $= 1$.

Proof. Follows from the convexity $uv \leq u \log(u) - u + e^v, u \geq 0, v \in \mathbb{R}$.

Alternatively, reduce by homogeneity to $\int f d\mu = 1$, and then use Jensen for the concave \log and the law $f d\mu$:

$$\int f g d\mu = \int f \log(f) d\mu + \int \log\left(\frac{e^g}{f}\right) f d\mu \leq \int f \log(f) d\mu + \log \int \frac{e^g}{f} f d\mu \leq \int f \log(f) d\mu.$$

\square

Replacing g such that $\int e^g d\mu = 1$ by $g - \log \int e^g d\mu$ without constraint on g , we get that relative entropy and log-Laplace transform are the Legendre transform of each other :

$$H(\nu | \mu) = \sup_g \left\{ \int g d\nu - \log \int e^g d\mu \right\} \quad \text{and} \quad \sup_\nu \left\{ \int g d\nu - H(\nu | \mu) \right\} = \log \int e^g d\mu.$$

Lemma 3.5. Tensorisation.

If $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ is a product probability measure on a product space $E = E_1 \times \cdots \times E_n$ then

$$\text{Ent}_\mu(f) \leq \sum_{i=1}^n \int \text{Ent}_{\mu_i}(f) d\mu \quad \text{for all } f \in L^1(\mu, E \rightarrow \mathbb{R}_+).$$

Proof. By induction on n , it suffices to consider the case $n = 2$. Let $g : E \rightarrow \mathbb{R}$ be such that $\int e^g d\mu = 1$. Then

$$g = g_1 + g_2 \quad \text{avec } g_1 := g - \log \int e^g d\mu_1 \quad \text{et } g_2 := \log \int e^g d\mu_1,$$

in such a way that $\int e^{g_1} d\mu_1 = 1$ and $\int e^{g_2} d\mu_2 = 1$. The variational formula of Lemma 3.4 for μ_i and g_i gives

$$\int f g_1 d\mu_1 + \int f g_2 d\mu_2 \leq \text{Ent}_{\mu_1}(f) + \text{Ent}_{\mu_2}(f), \quad \text{hence } \int f g d\mu \leq \int \text{Ent}_{\mu_1}(f) d\mu_1 + \int \text{Ent}_{\mu_2}(f) d\mu_2,$$

and it remains to use the variational formula of Lemma 3.4 this time for μ and g . \square

3.3 Log-Sobolev for Gaussian from two-points space via tensorization and CLT

Theorem 3.6. Logarithmic Sobolev inequality (LSI) for the Gaussian.

For all $n \geq 1$, denoting $\gamma_n := \mathcal{N}(0, I_n) = \gamma_1^{\otimes n}$, we have, for all $f \in L^2(\gamma^n) \cap \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$, in $[0, +\infty]$:

$$\text{Ent}_{\gamma^n}(f^2) \leq 2 \int |\nabla f|^2 d\gamma^n.$$

Moreover the constant 2 is optimal in the sense that equality is achieved for $f^2(x) = e^{\langle \lambda, x \rangle}$, $\lambda \in \mathbb{R}^n$.

- By analogy with classical Sobolev inequalities we can rewrite LSI as

$$\int f^2 \log(f^2) d\gamma^n \leq \int f^2 d\gamma^n \log \int f^2 d\gamma^n + 2 \int |\nabla f|^2 d\gamma^n,$$

stating that $f^2 \log(f^2) \in L^1(\gamma^n)$ as soon as $f^2 \in L^1(\gamma^n)$ and $|\nabla f|^2 \in L^1(\gamma^n)$.

- By an affine change of variable we get that $\mathcal{N}(m, \Sigma)$ satisfies an ISL with constant $\|\Sigma\|_{\text{op}}^2$.
- The linearization of LSI via $f^2 = (1 + \varepsilon g)^2$ gives a Poincaré inequality of constant 1 :

$$\text{Var}_{\gamma^n}(f) := \int f^2 d\gamma^n - \left(\int f d\gamma^n \right)^2 \leq \int |\nabla f|^2 d\gamma^n.$$

Proof following Gross and Bobkov. The idea is to start from the two-points space, tensorize to the cube, and then use the CLT. Namely, let us consider the uniform law $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$ on $\{-1, 1\}$. Then

$$\text{Ent}_\nu(g^2) \leq \frac{(g(1) - g(-1))^2}{2} \quad \text{for all } g : \{-1, 1\} \rightarrow \mathbb{R}.$$

We can assume without loss of generality that $g \geq 0$, and by homogeneity that $g(1)^2 + g(-1)^2 = 2$, which reduces the inequality to the optimal univariate inequality (checkable by direct calculus, equality achieved for $u = 1$)

$$u \log(u) + (2 - u) \log(2 - u) \leq (\sqrt{u} - \sqrt{2 - u})^2, \quad 0 \leq u \leq 2.$$

Now, let us take $f \in \mathcal{C}_c^2(\mathbb{R}, \mathbb{R})$ and let us define $g : \{-1, 1\}^n \rightarrow \mathbb{R}$ as

$$g(x_1, \dots, x_n) := f\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right).$$

Let $\mu := \nu^{\otimes n}$ be the uniform law on the cube $\{-1, 1\}^n$. By tensorisation (Lemma 3.5) and the inequality on $\{-1, 1\}$,

$$\text{Ent}_\mu(g^2) \leq \frac{1}{2} \int \sum_{i=1}^n (g(x^{i,+}) - g(x^{i,-}))^2 d\mu$$

where $x_j^{i,\pm} := x_j$ if $j \neq i$ and $:= \pm 1$ if $j = i$. A Taylor formula at order 1 for f at $\frac{x_1 + \dots + x_n}{\sqrt{n}}$ gives

$$g(x^{i,+}) - g(x^{i,-}) = \frac{2}{\sqrt{n}} f' \left(\frac{x_1 + \dots + x_n}{\sqrt{n}} \right) + o \left(\frac{1}{\sqrt{n}} \right)$$

with an o uniform in x since f is \mathcal{C}_c^2 and thus with bounded second derivative. Therefore, thanks to the CLT,

$$\text{Ent}_{\gamma_1}(f^2) \leq 2 \int f'^2 d\gamma^1.$$

We can weaken the conditions on f by approximation arguments.

We can generalize to $\gamma^n = (\gamma^1)^{\otimes n}$ for all $n \geq 1$ by using tensorization again! \square

- **Stability of LSI by tensorization or dimension free statements** : if μ, ν satisfy to LSI with constants c_μ and c_ν then $\mu \otimes \nu$ satisfies to LSI with constant $\max(c_\mu, c_\nu)$. In particular if μ satisfies LSI with constant c then $\mu^{\otimes N}$ satisfies to LSI with same constant c for all N . The constant depend on the class of test functions. The tensorization works if the class of test functions as well as the LHS are both stable by tensorization.
- **Stability by Lipschitz deformation**. If μ satisfies LSI with constant c and then its image with a map F satisfies LSI with constant $c \|F\|_{\text{Lip}}^2$. In particular Uniform $([0, 1])$ satisfies LSI, and LSI is stable by convolution.
- **Optimal transportation**. Caffarelli showed using the Monge–Ampère equation and the maximum principle that the Bakry–Émery condition implies that μ is the image of γ_n with F such that $\|F\|_{\text{Lip}} \leq \sigma$, leading to LSI via Lipschitz deformation from the Gaussian case. On the other hand, Cordero–Erasquin used Monge–Ampère to get LSI directly in this case, still via Monge–Ampère and an exploit of convexity.
- There is also a stability by bounded perturbation on V , due to Holley–Stroock. This was generalized by Bodineau–Helffer to V convex + bounded. Generalized by Zegarlinski to spin systems with exponential decay of correlations. Generalized by Bauerschmidt–Bodineau recently, in the spirit of high dimensional convexification...
- **Tails beyond Gaussians**. The probability measure $\frac{1}{Z_\alpha} e^{-|x|^\alpha} dx$ on \mathbb{R} , $\alpha > 0$, $Z_\alpha := \int_{\mathbb{R}} e^{-|x|^\alpha} dx < \infty$, satisfies LSI iff $\alpha \geq 2$, and a Poincaré inequality iff $\alpha \geq 1$. The Gaussian corresponds to the critical case $\alpha = 2$.

4 Log-Sobolev and concentration of measure

Theorem 4.1. LSI \Rightarrow sub-Gaussian Laplace transform of Lipschitz functions.

If $\mu \in \mathcal{P}(\mathbb{R}^n)$ satisfies to LSI with constant c :

$$\exists c \in \mathbb{R}_+, \forall f \in L^2(\mu) \cap \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}), \text{Ent}_\mu(f^2) \leq c \int |\nabla f|^2 d\mu.$$

then Lipschitz functions have sub-Gaussian Laplace transform :

$$\forall f : \mathbb{R}^n \rightarrow \mathbb{R} \text{ Lipschitz and in } L^1(\mu), \forall \theta \in \mathbb{R}, L(\theta) := \int \exp(\theta f) d\mu \leq \exp \left(\frac{\theta^2 c}{4} \|f\|_{\text{Lip}}^2 + \theta \int f d\mu \right).$$

Proof following Herbst. First of all we reduce to f bounded, \mathcal{C}^∞ , centered for μ , $\|f\|_{\text{Lip}} = 1$, and $\theta > 0$.

Now, for all $\theta > 0$, the LSI with $e^{\theta f}$ instead of f^2 gives, via $|\nabla e^{\theta f}| = |\theta \nabla f| e^{\theta f}$ and $\| |\nabla f| \|_\infty = \|f\|_{\text{Lip}} \leq 1$, that

$$\theta L'(\theta) - L(\theta) \log L(\theta) \leq \frac{c}{4} \theta^2 L(\theta), \quad \text{in other words } K' \leq \frac{c}{4} \text{ where } K(\theta) := \frac{1}{\theta} \log L(\theta).$$

The result follows from $K(0) = (\log L)'(0) = L'(0)/L(0) = \mu(f)$, which comes from $L(0) = 1$ and $L'(0) = \mu(f)$. \square

Corollary 4.2. LSI \Rightarrow Sub-Gaussian concentration for Lipschitz functions.

If $\mu \in \mathcal{P}(\mathbb{R}^n)$ satisfies to LSI of constant c as in Theorem 4.1,

then for all $X \sim \mu$, $r \geq 0$, and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in $L^1(\mu)$,

$$\mathbb{P}\left(\left|f(X) - \mathbb{E}(f(X))\right| \geq r\right) \leq 2 \exp\left(-\frac{r^2}{c\|f\|_{\text{Lip}}^2}\right).$$

More generally, if X_1, \dots, X_N , $N \geq 1$, are i.i.d. of law μ , then

$$\mathbb{P}\left(\left|\frac{f(X_1) + \dots + f(X_N)}{N} - \mathbb{E}(f(X_1))\right| \geq r\right) \leq 2 \exp\left(-\frac{Nr^2}{c\|f\|_{\text{Lip}}^2}\right).$$

- Dimension free rewrite :

$$\mathbb{P}\left(\sqrt{N}\left|\frac{f(X_1) + \dots + f(X_N)}{N} - \mathbb{E}(f(X_1))\right| \geq r\right) \leq 2 \exp\left(-\frac{r^2}{c\|f\|_{\text{Lip}}^2}\right).$$

- A consequence is the exponential integrability for the square of $Y := f(X) - \mathbb{E}(Y)$:

$$\mathbb{E}(e^{\theta Y^2}) = \theta \int_0^\infty r e^{\theta r^2} \mathbb{P}(|Y| \geq r) dr < \infty \quad \text{as soon as } \theta < \frac{1}{c\|f\|_{\text{Lip}}^2}.$$

Proof. For the first, we reduce to $\|f\|_{\text{Lip}} = 1$ and $\mu(f) = \int f d\mu = 0$ pas translation et dilatation, then for all $r \geq 0$ and $\theta > 0$, the Markov inequality and Theorem 4.1 give

$$\mu(f \geq r) = \mu(e^{\theta f} \geq e^{\theta r}) \leq e^{-\theta r} \int e^{\theta f} d\mu \leq e^{-\theta r + \frac{c}{4}\theta^2} \leq e^{-\frac{r^2}{c}},$$

where the last inequality comes from the optimal choice $\theta = 2r/c$. By using the result on $\pm f$ we get

$$\mu\left(\left|f - \int f d\mu\right| \geq r\right) \leq 2 \exp\left(-\frac{r^2}{2\|f\|_{\text{Lip}}^2}\right).$$

For the second inequality, we observe that $x \in (\mathbb{R}^n)^N \mapsto F(x) := \frac{1}{N}(f(x_1) + \dots + f(x_N))$ is Lipschitz with

$$\|F\|_{\text{Lip}} \leq \frac{\|f\|_{\text{Lip}}}{N} \sup_{x \neq y} \frac{\sum_{i=1}^N |x_i - y_i|}{\sqrt{\sum_{i=1}^N |x_i - y_i|^2}} \leq \frac{\|f\|_{\text{Lip}}}{\sqrt{N}}.$$

Moreover $\mathbb{E}(F(X_1, \dots, X_N)) = \mathbb{E}(f(X_1))$. Furthermore $(X_1, \dots, X_N) \sim \mu^{\otimes N}$ satisfies LSI with same constant $2c$ (dimension free : does not depend on N), thanks to the tensorization method used for proving Theorem 3.6. \square

- Unstability by tensor product of sub-Gaussianness of Laplace transform of Lipschitz functions and sub-Gaussian concentration, hence the usefulness of LSI when it holds!

4.1 Wigner Ensembles

Let $S := (S_{ij})_{1 \leq i, j \leq n}$ be an $n \times n$ real symmetric random matrix, $n \geq 1$.

Let $c_{ij} \in [0, +\infty]$ be the LSI constant of the law of S_{ij} (sparsity: take $c_{ij} = 0$ if S_{ij} is constant (possibly $\equiv 0$)).

Then for all $f : \mathbb{R} \rightarrow \mathbb{R}$ and all $r \geq 0$,

$$\mathbb{P}\left(\left|\text{Tr}_n f\left(\frac{S}{\sqrt{n}}\right) - \mathbb{E}\text{Tr}_n f\left(\frac{S}{\sqrt{n}}\right)\right| \geq r\right) \leq 2 \exp\left(-\frac{n^2 r^2}{\|f\|_{\text{Lip}}^2 \max_{ij} c_{ij}}\right).$$

LSI tensorization and spectrum of a symmetric matrix is a Lipschitz wrt its entries (Weyl inequalities) :

$$|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_{\text{op}}.$$

Special case : if S is Gaussian, say GOE, we can use the Gaussian LSI and the Lipschitz stability.

4.2 Beta-Ensembles

Let us consider the probability measure μ on \mathbb{R}^n given by

$$\frac{1}{Z_n} \prod_{i=1}^n e^{-\sum_{i=1}^n U(x_i)} \prod_{i<j} (x_i - x_j)^\beta \mathbf{1}_{x_1 \leq \dots \leq x_n} = \frac{1}{Z_n} e^{-\left(\sum_{i=1}^n U(x_i) + \beta \sum_{i<j} \log \frac{1}{x_i - x_j}\right)} \mathbf{1}_{x_1 \leq \dots \leq x_n}$$

where $U: \mathbb{R} \rightarrow \mathbb{R}$ is such that $U(x) = C(x) + \frac{1}{2\sigma^2} \|x\|^2$, C is \mathcal{C}^2 and convex and where $\sigma, \beta > 0$. Satisfies LSI with constant $2\sigma^2$, by Bakry–Émery or Caffarelli thanks to the convexity of

$$(x_1, \dots, x_n) \mapsto \sum_{i=1}^n C(x_i) - \beta \sum_{i<j} \log(x_i - x_j).$$

Draft

References (standard and personal)

- [1] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO et G. SCHEFFER : *Sur les inégalités de Sobolev logarithmiques*, vol. 10 de *Panor. Synth.* Paris: Société Mathématique de France, 2000.
- [2] D. BAKRY, I. GENTIL et M. LEDOUX : *Analysis and geometry of Markov diffusion operators*, vol. 348 de *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [3] J. BOURSIER, D. CHAFAÏ et C. LABBÉ : Universal cutoff for Dyson Ornstein Uhlenbeck process. *Probab. Theory Related Fields*, 185(1-2):449–512, 2023.
- [4] D. CHAFAÏ : From Boltzmann to random matrices and beyond. *Ann. Fac. Sci. Toulouse, Math.* (6), 24(4):641–689, 2015.
- [5] D. CORDERO-ERAUSQUIN : Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.*, 161(3):257–269, 2002.
- [6] A. GUIONNET et O. ZEITOUNI : Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136, 2000.
- [7] M. LEDOUX : *The concentration of measure phenomenon*, vol. 89 de *Math. Surv. Monogr.* Providence, RI: American Mathematical Society (AMS), 2001.
- [8] M. LEDOUX : More than fifteen proofs of the logarithmic Sobolev inequality. <http://perso.math.univ-toulouse.fr/ledoux/files/2023/03/15.pdf>, 2015.
- [9] C. VILLANI : *Topics in optimal transportation*, vol. 58 de *Graduate Studies in Mathematics*. AMS, 2003.