Concentration of Measure: Logarithmic Sobolev Inequalities

Aukosh Jagannath

September 15, 2013

Abstract

In this note we’ll describe the basic tools to understand Log Sobolev Inequalities and their relation to concentration. We will then briefly discuss applications.

1 Introduction

In the following the spaces we will be working with will be \((M^n, g, \mu(dx))\) where either \(M^n = \mathbb{R}^n\) or is a compact riemannian manifold and \(dx\) is either the lebesgue measure or the normalized riemannian metric and

\[
\mu(dx) = \frac{e^{-\Phi(x)}}{Z} dx
\]

where we assume \(\Phi\) is smooth.

For \(\Phi\) we define the operator

\[
L_\Phi = \Delta - < \nabla \Phi, \nabla \cdot >
\]

acting on \(C^\infty(M) \subset D(L_\Phi)\) where the terms are to be interpreted in the obvious way in the manifold setting. I’ll drop out the \(\Phi\) in the subscript from now on.

Definition 1. Let \(\Gamma_n(f, g)\) be defined as follows, \(\Gamma_0(f, g) = fg\), and

\[
\Gamma_n(f, g) = L\Gamma_{n-1}(f, g) - \Gamma_{n-1}(f, Lg) - \Gamma_{n-1}(g, Lf)
\]

In particular, when \(n = 1\), this is called the carré du champs operator and when \(n = 2\) this is called the carré du champs itéré operator.

With this we then can define LSI

Definition 2. A space \((M, g, \mu)\) as above admits a Logarithmic Sobolev Inequality if there is a \(c > 0\) such that

\[
\text{Ent}_\mu(f^2) := \int f^2 \log \left( \frac{f^2}{\int f^2 d\mu} \right) d\mu \leq 2c \int \Gamma_1(f, f) d\mu
\]

The main point of the LSI for us will be a consequence of the Herbst argument.

Theorem 3. (Gross-Herbst) Let \((M, g, \mu)\) satisfy and LSI with constant \(c\). Then for all bounded 1-lipschitz functions

\[
\mathbb{E}e^{\lambda(f - \mathbb{E}f)} \leq e^{\frac{\lambda^2}{2c}}
\]

In particular, the space admits subgaussian concentration

\[
\alpha_{(M, g, \mu)}(r) \leq e^{-\frac{r^2}{2c}}
\]
Proof. Suppose that we have the log sobolev inequality, assume $f$ is differentiable and bounded and mean zero. Let \( \Lambda(\lambda) = \log \mathbb{E} e^{\lambda f} \) then, letting $F^2 = e^{\lambda f}$

\[
\frac{d}{d\lambda} \Lambda(\lambda) = \frac{1}{\lambda} \left( \frac{1}{\mathbb{E} e^{\lambda f}} \mathbb{E} (f - \mathbb{E} f) e^{\lambda f} \right) - \frac{\log \mathbb{E}...}{\lambda^2}
= \frac{1}{\lambda^2 \mathbb{E} F^2} \left( \mathbb{E} (f - \mathbb{E} f) e^{\lambda f} \right) - \frac{\mathbb{E} (f - \mathbb{E} f) e^{\lambda f} \log \mathbb{E} F^2}{\lambda^2 \mathbb{E} F^2}
= \frac{\mathbb{E} (f^2) - \mathbb{E}^2 (f) \log \mathbb{E} F^2}{\lambda^2 \mathbb{E} F^2}
= \frac{L_{SI} \mathbb{E} F^2}{\lambda^2 \mathbb{E} F^2}
= \frac{\mathbb{E} e^{\lambda f} - \mathbb{E} e^{\lambda f} e^{\lambda f} \log \mathbb{E} F^2}{\lambda^2 \mathbb{E} F^2}
\]

since $f$ is mean 0, the left hand side is 0 at 0. Integrating the inequality and exponentiating gives

\[
\mathbb{E} e^{\Lambda(f)} \leq e^{\frac{\lambda^2}{2}}.
\]

Subtracting the mean mollifying extends the result to the full class. \( \square \)

Remark 4. Recall that by standard arguments this extends to all lipschitz functions, which you can show must be in $L_1(\mu)$ by subgaussian concentration.

Logarithmic sobolev inequalities have several non trivial reasons why they are interesting. Perhaps the simplest is the dimension free nature of the inequality under $\ell_2$ products. That is, if we consider $(M_i, g_i \mu_i)$ and we take their measure product with the metric given by the product metric on riemannian manifolds, which is an $\ell_2$ product, then we get that the new product gets a LSI with the optimal constant the smallest of the others.

Theorem 5. (Efron-stein)

\[
\text{Ent}(f) \leq \sum_{i=1}^{N} \int \text{Ent}_{\mu_i} (f) d\mu^{\otimes k=1}_{i=1}
\]

Proof. Note that

\[
\text{Ent}(f) = \sup \left\{ \int f g, \int e^g d\mu = 1 \right\}
\]

take $g$ on the product and let $g^i(x_i, \ldots, x_n) = \log \left( \frac{e^{g(x_i, \ldots, x_n) d\mu_i}}{e^{g(x_i, \ldots, x_n) d\mu_i}} \right)$ so that by jensens inequality,

\[
g \leq \sum_i g^i
\]

thus

\[
\int f g \leq \sum_i \int f g^i
\]

since $\int e^{g_i} = 1$, we see that maximizing and passing the sup through the sum gives the result. \( \square \)

2 The Bakry-Emery condition

Logarithmic sobolev inequalities hold in a variety of spaces. A simple condition that holds in a very large class of spaces is the Bakry-Emery condition.
Definition 6. A space satisfies the Bakry-Emery condition if
\[ \Gamma_2(f, f) \geq \frac{1}{c} \Gamma_1(f, f) \]
for some \( c > 0 \)

Remark 7. On \( M = \mathbb{R}^n \), We will see below that this condition is equivalent to \( \text{Hess} \Phi \geq c \text{Id} \)

We can relate the above operators to more classical objects. Recall the dirichlet form

Definition 8. The dirichlet form for \( L \) is given by
\[ \mathcal{E}(f, g) = -(f, Lg) \]

We can then show the following formulae.

Proposition 9. The carré du champ operator satisfies
\[ \Gamma_1(f, f) = g(\nabla f, \nabla f). \]
and we have the Bochner-Bakry-Emery formula
\[ \Gamma_2(f, f) = \langle \text{Hess}(f), \text{Hess}(f) \rangle + (\text{Ric} + \text{Hess}\Phi)(\nabla f, \nabla f). \]
Furthermore, we have the integration by parts formula
\[ \mathcal{E}(f, g) = \int \Gamma_1(f, g) d\mu \]

With this in hand we can now show the main theorem in this section.

Theorem 10. If a space satisfies the BE condition with constant \( c \) then it has an LSI with constant \( c \).

Lemma 11. A space satisfies the BE condition with constant \( c \) iff the carré du champ operator satisfies
\[ \Gamma_1(P_t f, P_t f) \leq e^{-\frac{c}{2}t} P_t \Gamma_1(f, f) \]

Proof. Let \( \psi(s) = P_s \Gamma_1(P_{t-s} f, P_{t-s} f) \) where \( f \in D(L) \) in fact we’ll assume that \( f \) is smooth and, if \( M^n = \mathbb{R}^n \) then it can be taken to be at most poly growth at infinity. Then
\[ \psi'(s) = L P_s \Gamma_1(P_{t-s} f, P_{t-s} f) - 2 P_s \Gamma_1(L P_{t-s} f, P_{t-s} f) \]
\[ =: P_s \Gamma_2(P_{t-s} f, P_{t-s} f) \]
\[ \leq \frac{1}{c} P_s \Gamma_1(P_{t-s} f, P_{t-s} f) \]
\[ = \frac{1}{c} \psi(s) \]
solving the differential inequality and plugging in \( s = t, 0 \) gives the result.

To go the other way, note that
\[ 0 \leq \left( e^{-\frac{c}{2}t} P_t \Gamma_1(f, f) - \Gamma_1(f, f) \right) /t - (\Gamma_1(P_t f, P_t f) - \Gamma_1(f, f)) /t \]
sending \( t \to 0 \)
\[ 0 \leq -\frac{2}{c} \Gamma_1(f, f) + LL \Gamma_1(f, f) - 2 (\Gamma_1(L f, f)) \]
\[ \square \]
Remark 12. Note that using the standard grown wall inequality argument we can already get subgaussian concentration without passing to the log-sobolev arguments.

Theorem 13. The BE condition implies Ergodicity (Assuming that your manifold is geodesically complete).

That is, $P_t f \overset{L^2}{\rightarrow} \mu f$ and almost surely

Proof. Let $f$ be some sufficiently regular function (see the previous proof), and let $f_t(x) = P_t f$, then if $\gamma(s)$ is a geodesic starting at $y$ and ending at $x$,

$$|f_t(x) - f_t(y)| = \int_0^1 \langle \nabla f_t(\gamma(s)), \dot{\gamma}(s) \rangle \, ds$$

$$\leq \int_0^1 \left( \int_0^1 \Gamma_1(f_t, f_t) \, ds \right)^{1/2} \, d(x, y)$$

$$\leq \exp \left( -\frac{t}{c} \left( \int_0^1 \Gamma_1(f, f) \, ds \right)^{1/2} \right) \, d(x, y).$$

assuming your manifold is cpt or that your function is lipschitz gives a lipschitz constant of $e^{-t/c}||f||_{Lip}$. sending $t \to \infty$ gives the almost sure convergence. since

$$\mu|P_t f|^{2+\epsilon} \leq \mu|f|^{2+\epsilon}$$

by jensens, (take $f \in C_0$ on $\mathbb{R}^n$) we have by uniform integrability and thus the result for convergence in $L_2$. We can extend this result further by obvious means. \qed

Remark 14. What this result tells us is that it gives us a continuum analogue of the glauber dynamics result, that is if we propose a sufficiently regular measure, then we have a dynamics to recover it as a stationary distribution.

With these results we can now prove the main result of the section

Proof. (Log-Sobolev) First note that by the above arguments and 19, if we assume $f \geq \epsilon$

$$\text{Ent}(f) = \int P_0 f \log P_0 f \, d\mu - \int P_\infty f \log P_\infty f \, d\mu$$

$$= -\int_0^\infty \left( \int P_s f \log P_s f \, d\mu \right)' \, ds$$

$$= -\int_0^\infty \left( \int LP_s f \log P_s f + LP_s f \, d\mu \right) \, ds$$

$$\overset{\text{Lip}}{=} -\int_0^\infty \left( \int LP_s f \log P_s f \right) \, ds$$

$$= \int_0^\infty \Gamma_1(P_s f, \log P_s f) \, d\mu \, ds.$$
Now

\[ \int \Gamma_1(P_s f, \log P_s f) d\mu \xrightarrow{\text{symm.}} \int \Gamma_1(f, P_s (\log(P_s f))) d\mu \]

\[ \leq e^{-\frac{3}{2}} \left( \int \frac{\Gamma_1(f,f)}{f} \right)^{1/2} \left( \int P_s f \Gamma_1(\log(P_s f), \log(P_s f)) d\mu \right)^{1/2} \]

\[ \leq e^{-\frac{3}{2}} \left( \int \frac{\Gamma_1(f,f)}{f} \right)^{1/2} \left( \int \Gamma_1(P_s f, \log P_s f) d\mu \right)^{1/2} \]

so that, replacing \( f \) with \( f^2 \)

\[ \text{Ent}(f^2) \leq \int_0^\infty e^{-\frac{3}{2}} \int \frac{\Gamma_1(f^2, f^2)}{f^2} d\mu \]

\[ = 4 \int_0^\infty e^{-\frac{3}{2}} \int \Gamma_1(f,f) d\mu ds \]

\[ = 2e \int \Gamma_1(f,f) d\mu \]

now again we can extend to the full class regularizing \( f^2 + \epsilon \)

\[ \square \]

3 Examples

**Proposition 15.** The space \((\mathbb{R}^n, \mu_\Phi)\) with \(\text{Hess}\Phi \geq cId\) where \(c \geq 0\) satisfies the Bakry-Emery condition with constant \(c\). In fact, if we have \(\text{Ric} + \text{Hess}\Phi \geq cg\) with then we have the BE conditions with constant \(c\).

**Proof.** Recall that \(\Gamma_2(f,f) = (\text{Hess}(f), \text{Hess}(f)) + (\text{Ric} + \text{Hess}\Phi)(\nabla f, \nabla f)\). To get the first, note that \(\text{Ric} = 0\) and so if we take \(f = \sum a_i x_i\) this condition is equivalent to \(\text{Hess}(\Phi) \geq \frac{1}{c}Id\). The second argument is clear. \(\square\)

**Example 16.** Gauss space \((\mathbb{R}^n, \gamma_n)\) admits a logarithmic Sobolev inequality with optimal constant \(c = 1\). To see this note that \(\Phi(x) = \frac{x^2}{2}\) so we get the above conditions with \(c = 1\).

**Example 17.** Manifolds with lower Ricci bounds and the normalized volume form have no hessian term so, \(\text{Ric} \geq cg\). Thus it satisfies the BE conditions with constant \(c\) and thus an LSI with constant \(c\). Note that for the classical homogenous spaces, the quotients of \(SO(n)\), we can show that \(\text{Ric} = \left(\frac{N-2}{4}\right)g\) so that we have subgaussian concentration.

4 Application: Random Matrices

LSI’s lend themselves even to correlated settings

**Theorem 18.** Consider \(d\mu = \frac{e^{-Ntr(V(x))} dvol}{Z}\) where \(V_N(x)\) is twice differentiable and strictly convex with \(V'' \geq c\) and \(dvol\) is the volume measure on the real symmetric \(N \times N\) matrices. Then the ensemble has LSI. In particular for all lipschitz \(f : \mathbb{R}^N \rightarrow \mathbb{R}\),

\[ P(|f(\lambda_1, \ldots, \lambda_N) - Ef| \geq r) \leq e^{-\frac{Nr^2}{2\|f\|_{Lip}^2}} \]
Proof. Working in coordinates one can see that $tr V$ is also twice differentiable and strictly convex with the same constant, giving us the LSI result. the map to the eigenvalues is lipschitz by the Hoffman-Weilandt lemma giving us the other result by the contraction principle.

Appendix

Proposition 19. A Markov Semigroup is symmetric (i.e. $(P_t f, g) = (f, P_t g)$) if and only if the infinitesimal operator is self adjoint

Proof. going from symmetry of the operator to selfadjointness is just the definition of the operator. To go the other way, recall that the key element of the proof of the Hille-Yosida theorem is the construction of the Yosida regularizers $L_\lambda = L\lambda(\lambda - L)^{-1} \in B(H)$ which will converge to $L$ as $\lambda \rightarrow \infty$ point wise on its domain, from there one goes about constructing the semigroup by using the exponential series for $L_\lambda$. use this to show symmetry on the domain of $D(L)$ and extend by density.