Improved log Sobolev coefficients for compact Lie groups

Tai Melcher
University of Virginia

joint work with E. Meckes and ongoing work with J. Wang

Southeastern Probability Conference
in honor of Elizabeth Meckes
18 May 2021

Log Sobolev inequalities

A metric space (X, d) equipped with a Borel probability measure ρ satisfies a log Sobolev inequality (LSI) with constant $C > 0$ if, for every locally Lipschitz function $f : X \to \mathbb{R}$,

$$\int (f^2 \log f^2) d\rho - \left(\int f^2 d\rho \right) \log \left(\int f^2 d\rho \right) \leq 2C \int |\nabla f|^2 d\rho,$$

where

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y, x)}.$$

We'll be focused on the setting of X a Riemannian manifold (M, g) equipped with the standard Riemannian distance $d = d_g$ and $\rho_t = \mathrm{Law}(B_t)$ where B_t is a Brownian motion on M.

Brownian motion on a Lie group

In particular, when $M = G$ is a Lie group, the Brownian motion may be expressed as the solution of the SDE

$$dB_t = B_t \circ dW_t$$

with $B_0 = e$, where W_t is BM on $\mathfrak{g} = \mathrm{Lie}(G)$.

As an illustrating example, we focus on the even more specific case that $G = U(N)$ is the $N \times N$ unitary matrices with Lie algebra $\mathfrak{u}(N)$, the skew-Hermitian matrices. For $A, B \in \mathfrak{u}(N)$, let

$$\langle A, B \rangle = N\mathrm{Tr}(B^* A)$$

and, identifying $T_I U(N) \cong \mathfrak{u}(N)$, we may extend this to a bi-invariant Riemannian metric on $U(N)$.

Brownian motion on $U(N)$

We may describe Brownian motion U_t^N on $U(N)$ as the solution to the SDE

$$\begin{aligned} dU_t^N &= U_t^N \circ dW_t^N \\ &= U_t^N \, dW_t^N - \frac{1}{2} U_t^N \, dt \end{aligned}$$

with $U_0^N = I_N$, where W_t is a standard Brownian motion on $\mathfrak{u}(N)$ (for example, take $\{\xi_k\}_{k=0}^{N^2-1}$ an onb of $\mathfrak{u}(N)$ wrt the given inner product and $W_t^N = \sum_{j=0}^{N^2-1} b_t^j \xi_j$, where the b_t^j are independent standard BMs on $\mathbb{R}$).

We're interested in understanding LSIs for $\mathrm{Law}(U_t^N)$.

Some motivation: The empirical spectral measure

A matrix $U \in U(N)$ has N eigenvalues on S^1, $e^{i\theta_1}, \dots, e^{i\theta_N}$.
The spectral measure of U is the probability measure on S^1

$$\mu_U := \frac{1}{N} \sum_{j=1}^{N} \delta_{e^{i\theta_j}}.$$

For each fixed $t > 0$, U_t is a random unitary matrix, and we denote its empirical spectral measure by $\mu_t^N := \mu_{U_t}$.

Spectrum of BM on $U(20)$

Click!

Some motivation: ESM $\to$ Biane's measure

Theorem (Biane, 1997)

There exists a deterministic probability measure ν_t on S^1 so that μ_t^N converges weakly almost surely to ν_t:
for all $f \in C(S^1)$,

$$\lim_{N\to\infty} \int_{S^1} f \, d\mu_t^N = \int_{S^1} f \, d\nu_t \text{ a.s.}$$

Rates of convergence of $\mu_t^N \to \nu$?

The equilibrium case

Theorem (Meckes–Meckes, PTRF 2013)

Let U^N be a Haar-distributed (uniform) random matrix in $U(N)$, and let ν be the uniform measure on S^1. Then

$$\mathbb{E}[W(\mu_{U^N}, \nu)] \leq C \frac{\log N}{N}$$

and with probability 1

$$W(\mu_{U^N}, \nu) \leq C \frac{\log N}{N}$$

for all sufficiently large N.

Proof relied on concentration of measure techniques.

Concentration via LSI

Let (X, d) denote a metric space with Borel probability measure ρ.

If (X, d, ρ) satisfies a LSI with constant $C > 0$, then for any L-Lipschitz $F : X \to \mathbb{R}$ with $\mathbb{E}|F| < \infty$

$$\rho\left(|F - \mathbb{E}F| \geq r\right) \leq 2e^{-r^2/L^2 C}.$$

LSI for heat kernel measure

When (X, d) is a Riemannian manifold (M, g, d_g), let ∇ and Δ denote the gradient and Laplace-Beltrami operators acting on $C^\infty(M)$. Let

$$P_t f(x) := e^{t\Delta/2} f(x) := \mathbb{E}[f(B_t^x)] = \int_M f \, d\rho_t^x$$

where $\{B_t^x\}_{t\geq 0}$ is a Brownian motion on M started at x.

(i) [Bakry–Émery, 1984] $\mathrm{Ric} \geq -k$ if and only if

$$|\nabla P_t f| \leq K(t) P_t |\nabla f|, \forall f \in C_c^\infty(M) \text{ and } t > 0, \qquad (*)$$

holds with $K(t) = e^{kt}$.

(ii) If (*) holds, then LSI is satisfied for ρ_t^x with coefficient

$$C(t) = \int_0^t K^2(s) ds.$$

LSI for heat kernel measure

(i) $\mathrm{Ric} \geq -k$ if and only if

$$|\nabla P_t f| \leq K(t) P_t |\nabla f|, \forall f \in C_c^\infty(M) \text{ and } t > 0, \qquad (*)$$

holds with $K(t) = e^{kt}$.

(ii) If (*) holds, then LSI is satisfied for ρ_t^x with coefficient

$$C(t) = \int_0^t K^2(s) ds.$$

(i) $\mathrm{Ric}_{U(N)} \geq 0 \implies$

$$|\nabla P_t f| \leq P_t |\nabla f|, \forall f \in C_c^\infty(U(N)) \text{ and } t > 0,$$

(ii) Thus LSI for $\mathrm{Law}(U_t)$ is satisfied with coefficient $C(t) = t$.

But we should be able to do better. . . .

Transformations of LSI

It's well-known that LSI behaves well under certain operations.

- (tensorization) Suppose that (X_i, d_i, ρ_i) satisfy LSI with constant $C_i < \infty$. Then for $X = X_1 \times \cdots \times X_n$ equipped with product distance and product probability measure $\rho := \rho_1 \otimes \cdots \otimes \rho_n$, (X, d, ρ) satisfies LSI with constant $C := \max_{1 \le i \le n} C_i$.
- (under Lipschitz mappings) Suppose that (X, d_X, ρ) satisfies LSI with constant C. Let $F : (X, d_X) \to (Y, d_Y)$ be an L-Lipschitz map. Then $(Y, d_Y, \rho \circ F^{-1})$ satisfies LSI with constant $L^2 C$.

We combine these tools and the known estimate from the curvature bound, with a smart coupling of Brownian motions, inspired by an analogous result for Haar measure on $U(N)$ [Meckes–Meckes, ECP 2013]

The coupling

Lemma (Meckes-M)

Let b^0 be an $\mathbb{R}$-valued Brownian motion and $z_t := e^{ib_t^0/N}$, and let V_t be a Brownian motion on $SU(N)$ issued from the identity. Then $z_t V_t$ is a Brownian motion on $U(N)$.

Proof. Note that z_t satisfies the SDE

$$dz_t = z_t \frac{idb_t^0}{N} - \frac{1}{2N^2} z_t \, dt.$$

Setting $Z_t := z_t I_N$

$$dZ_t = Z_t \, db_t - \frac{1}{2N^2} Z_t dt$$

where $b_t = b_t^0 \xi_0$ with $\xi_0 = iI_N/N$.

The coupling

Let $\beta = \{\xi_j\}_{j=1}^{N^2-1}$ be an onb of $\mathfrak{su}(N)$, and let $\{b_t^j\}_{j=1}^{N^2-1}$ be independent $\mathbb{R}$-valued BM. Then

$$\tilde{W}_t = \sum_{j=1}^{N^2-1} b_t^j \xi_j$$

is a BM on $\mathfrak{su}(N)$, and V_t satisfies the SDE

$$\begin{aligned} dV_t &= V_t \circ d\tilde{W}_t \\ &= V_t\, d\tilde{W}_t + \frac{1}{2} V_t \sum_{\xi\in\beta} \xi^2\, dt \\ &= V_t\, d\tilde{W}_t - \left(\frac{N^2-1}{2N^2}\right) V_t\, dt. \end{aligned}$$

The coupling

Now, $\{\xi_j\}_{j=0}^{N^2-1}$ is an onb of $\mathfrak{u}(N)$, and $z_t V_t = Z_t V_t$ satisfies

$$
\begin{aligned}
d(Z_t V_t) &= \left(Z_t db_t - \frac{1}{2N^2} Z_t \, dt \right) V_t + Z_t \left(V_t \, d\tilde{W}_t - \left(\frac{N^2-1}{2N^2} \right) V_t \right) \\
&= Z_t V_t \, (db_t + d\tilde{W}_t) - \frac{1}{2} Z_t V_t \, dt.
\end{aligned}
$$

Since $W_t = b_t + \tilde{W}_t$ is a BM on $\mathfrak{u}(N)$, and that BM on $U(N)$ is the solution to

$$
dU_t = U_t dW_t - \frac{1}{2} U_t dt,
$$

this implies that $z_t V_t \in SU(N) \rtimes S^1 \simeq U(N)$ is a BM on $U(N)$. ■

Improving LSI on $U(N)$

- With this scaling of the metric $\mathrm{Ric}_{SU(N)} = \frac{1}{2}$. Thus LSI holds on $SU(N)$ for $\mathrm{Law}(V_t)$ with coefficent

$$C(t) = 4(1 - e^{-t/4}).$$

- Elizabeth's LEMMA: for the heat semi-group on S^1,

$$|(P_t f)'| \leq \frac{e^{-(t-a)/4}}{1 - e^{-(t-a)/4}} P_t |f'|.$$

  This proof was based on [Saloff-Coste 1994] uniform estimates on the distance from the heat kernel on S^1 to 1.

  Combining this with $\mathrm{Ric}_{S^1} \geq 0$ shows that the LSI holds on S^1 for $\mathrm{Law}(e^{ib_t})$ with coefficient $C(t) = \min\{t, C\}$.

Improving LSI on $U(N)$

- Define $F : S^1 \times SU(N) \to U(N)$

$$F(e^{i\theta}, V) = e^{i\theta} V.$$

Then F is Lipschitz and

$$F(e^{ib_{t/N^2}}, V_t) \stackrel{d}{=} z_t V_t \stackrel{d}{=} U_t.$$

Theorem (Meckes–M)

The LSI holds for the heat kernel measure $\mathrm{Law}(U_t)$ on $U(N)$ with coefficient $C(t) = C' \min\{t, C\}$.

Application to esm $\rightarrow$ Biane's measure

Theorem (Meckes–M, 2018)

With probability one for N sufficiently large

$$\sup_{0 \leq t \leq T} W_1(\mu_t^N, \nu_t) \leq c \frac{T^{2/5} \log N}{N^{2/5}}.$$

- previously known convergence rates were for moments of the ensemble-averaged spectral measure [Collins–Dahlqvist–Kemp 2016]
- paper borrowed multiple techniques from Meckes–Meckes toolkit
- the proof of concentration that appeared in the paper didn't go through LSI

General compact Lie groups as products

Proposition (Cheeger–Ebin 1975, Milnor 1976)

A simply connected Lie group G which admits a bi-invariant metric is the product of a (compact) group with strictly positive Ricci curvature and a group of curvature 0.

Let

$$\mathfrak{z} := \{x \in \mathfrak{g} : [x, y] = 0 \text{ for all } y \in \mathfrak{g}\}$$

which is an ideal of $\mathfrak{g}$. If G has a bi-invariant metric and I is any ideal of $\mathfrak{g}$ then $I^\perp$ is also an ideal since

$$0 = \langle I, I^\perp \rangle = \langle [x, I], I^\perp \rangle = \langle I, [x, I^\perp] \rangle.$$

Thus, $\mathfrak{g}$ splits as $\mathfrak{g} = \mathfrak{z} \oplus \mathfrak{h}$ where $\mathfrak{h} = \mathfrak{z}^\perp$. Let $G = Z \times H$ be the corresponding splitting of G.

The coupling on general compact Lie groups

Let $Z(t)$ and $H(t)$ be Brownian motions on Z and H:

$$dZ(t) = Z(t)\sum_{k=1}^{m} z_k \circ dB_k(t), \quad dH(t) = H(t)\sum_{j=1}^{n} h_j \circ dW_j(t)$$

where $\{B_k\}, \{W_j\}$ are independent standard BMs and $\{z_k\}$ is an onb of $\mathfrak{z}$ and $\{h_j\}$ an onb of $\mathfrak{h}$. Then

$$d(Z(t)H(t)) = Z(t)\sum_{i=1}^{m} z_i \circ dB_i(t)H(t) + Z(t)H(t)\sum_{j=1}^{n} h_j \circ dW_j(t).$$

$\implies Z(t)H(t)$ is a Brownian motion on G.

Improved LSI on general compact Lie groups

Theorem (Meckes–M–Wang, 21+)

For $t > 0$, the heat kernel measure on G satisfies the LSI with coefficient $C(t) = C' \min\{t, C\}$.

Follows essentially as before from the tensorization property of the LSI and the coefficient for the heat kernel measure on $Z \sim (\mathbb{S}^1)^{\otimes m}$.