Deficit bounds for log Sobolev inequality

Daesung Kim

Department of Mathematics
University of Illinois at Urbana–Champaign

March 1, 2021

Logarithmic Sobolev inequality

For two probability measures μ and ν with $\mu \ll \nu$,

$$\mathrm{I}(\mu|\nu) = \int \left|\nabla \left(\log \frac{d\mu}{d\nu}\right)\right|^2 d\mu, \qquad \text{(Fisher information)}$$

$$\mathrm{H}(\mu|\nu) = \int \log \frac{d\mu}{d\nu}\, d\mu. \qquad \text{(relative entropy)}$$

Let $d\nu = d\gamma = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} dx$ and $d\mu = f d\gamma$, then

$$\mathrm{I}(f) = \mathrm{I}(\mu|\nu) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}\, d\gamma,$$

$$\mathrm{H}(f) = \mathrm{H}(\mu|\nu) = \int_{\mathbb{R}^n} f \log f\, d\gamma.$$

The (Gaussian) logarithmic Sobolev inequality states

$$\frac{1}{2}\mathrm{I}(f) \geq \mathrm{H}(f).$$

Logarithmic Sobolev inequality

There are several equivalent forms of the LSI:

$$\begin{aligned}
\frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}\,d\gamma &\geq \int_{\mathbb{R}^n} f\log f\,d\gamma,\\
2\int_{\mathbb{R}^n} |\nabla g|^2\,d\gamma &\geq \int_{\mathbb{R}^n} g^2\log g^2\,d\gamma,\\
\frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla h|^2}{h}\,dx &\geq \int_{\mathbb{R}^n} h\log h\,dx + \frac{n}{2}\log(2\pi e^2).\\
\frac{1}{2\pi}\int |\nabla f|^2\,dm &\geq \int |f|^2\log|f|^2\,dm \quad (dm = 2^{\frac{n}{2}}e^{-2\pi|x|^2}\,dx).
\end{aligned}$$

The constant $\frac{1}{2}$ is sharp and dimension-free.

Equality holds if and only if $f = e^{b\cdot x - \frac{1}{2}|b|^2}$.

Among centered probabiilty measures $f d\gamma$ (that is, $\int f\,d\gamma = 0$), $f = 1$ is the unique optimizer.

Gross (1975): LSI and Hypercontractivity

Gross (1975) proved that the LSI is equivalent to the hypercontractivity.

The Ornstein–Uhlenbeck semigroup $P_t f(x) = \int f(e^{-t}x + \sqrt{1-e^{-t}}y)\, d\gamma$ is hypercontractive, that is, for $1 < p < q < \infty$,

$$\|P_t f\|_q \le \|f\|_p$$

for $t \ge T(p,q) = \frac{1}{2}\log\frac{q-1}{p-1}$.

If $q(t) = 1 + (p-1)e^{2t}$, then $F(t) = \|P_t f\|_{q(t)}^{q(t)}$ is non-decreasing in t by the hypercontractivity.

$F'(0) \le 0$ with $p = 2$ yields the LSI.

Gross (1975): Two-point Process

Let $X = \{1, -1\}$ and μ be a probability measure on X such that $\mu(\{-1\}) = \mu(\{1\}) = \frac{1}{2}$.

For $f : X \to \mathbb{R}$ with $\int_X f^2\, d\mu = 1$, we have the LSI

$$\frac{1}{2}\int_X |Df|^2\, d\mu = \frac{1}{2}\int_X |f(1) - f(-1)|^2\, d\mu \geq \int_X f^2 \log f^2\, d\mu.$$

By tensorization, one can extend it to $X^n = \{1, -1\}^n$

$$\frac{1}{2}\int_{X^n} \sum_{i=1}^{n} |D_i f|^2\, d\mu \geq \int_{X^n} f^2 \log f^2\, d\mu.$$

The Gaussian LSI on $\mathbb{R}^n$ follows from the central limit theorem.

Carlen (1991): Bechner–Hirschman uncertainty principle

Let $g(x) = 2^{\frac{n}{4}} e^{-\pi |x|^2}$, $dm = g(x)^2 dx$.

Bechner–Hirschman uncertainty principle states that

$$S(|h|^2) + S(|\widehat{h}|^2) \geq n(1 - \log 2)$$

where $\widehat{h}$ is the Fourier transfrom of h and $S(\rho) = -\int \rho \log \rho \, dx$.

The Fourier–Wiener transform $\mathcal{W}$ on $L^2(dm)$ is defined by $\mathcal{W}f = \widehat{(fg)}/g$.

Carlen derived the following deficit estimate from the entropic uncertainty principle, to characterize the cases of equality in the LSI.

$$\frac{1}{2\pi} \int |\nabla f|^2 \, dm - \int |f|^2 \log |f|^2 \, dm \geq \int |\mathcal{W}f|^2 \log |\mathcal{W}f|^2 \, dm \geq 0.$$

Ledoux (1992): Along Ornstein–Uhlenbeck semigroup

Let P_t be the Ornstein–Uhlenbeck semigroup defined by

$$P_t f(x) = \int f(e^{-t}x + \sqrt{1 - e^{-2t}}y)\, d\gamma(y).$$

Consider

$$F(t) = \mathrm{I}(P_t f) = \int \frac{|\nabla P_t f|^2}{P_t f}\, d\gamma,$$

then $\int_0^\infty F(t)\, dt = \mathrm{H}(f)$, $F(0) = \mathrm{I}(f)$, and

$$F'(t) = -2F(t) - 2e^{-4t} \int \frac{1}{|P_t f|^3} \sum_{i,j=1}^{n} (P_t f_i P_t f_j - P_t f P_t(f_i f_j))^2\, d\gamma.$$

Integrating over t, we get

$$\frac{1}{2}\mathrm{I}(f) - \mathrm{H}(f) = \int_0^\infty e^{-4t} \int \frac{1}{|P_t f|^3} \sum_{i,j=1}^{n} (P_t f_i P_t f_j - P_t f P_t(f_i f_j))^2\, d\gamma dt \geq 0.$$

Cordero-Erausquin (2002): Optimal transportation

Let μ and ν be Borel probability measures on $\mathbb{R}^n$.

We say that a map $T : \mathbb{R}^n \to \mathbb{R}^n$ pushes μ forward to ν if $\nu(B) = \mu(T^{-1}(B))$ for every Borel set $B \subset \mathbb{R}^n$.

Brenier and McCann showed that there exists a convex function φ such that $T = \nabla\varphi$ pushes μ forward to ν.

Let $T = \nabla\varphi$ be the Breiner map between $f d\gamma$ and $d\gamma$ and $\theta(x) = \varphi(x) - \frac{|x|^2}{2}$, then

$$\frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}\, d\gamma - \int_{\mathbb{R}^n} f \log f \, d\gamma$$
$$\geq \frac{1}{2}\int |\nabla(\log f) + \nabla\theta|^2 f d\gamma + \int (\Delta_A \theta - \log\det(I + \operatorname{Hess}\theta))\, f d\mu.$$

Using $\log(1+t) \leq t$, we have

$$\int (\Delta_A \theta - \log\det(I + \operatorname{Hess}\theta))\, f d\mu \geq 0.$$

Stability of the LSI

We call $\delta(f) = \frac{1}{2}\mathrm{I}(f) - \mathrm{H}(f)$ the LSI deficit.

Stability question

If $\delta(f) \to 0$, then does f converge to an optimizer? If we restrict to centered probability measures, $f \to 1$? In what sense?

Non-quantitative stability results

For a distance d on a class of centered probability measures X,

$$\delta(f) \to 0 \qquad \Longrightarrow \qquad \mathrm{d}(f d\gamma, d\gamma) \to 0$$

Quantitative stability results

For a distance d on a class of centered probability measures X,

$$\delta(f) \geq \Phi(\mathrm{d}(f d\gamma, d\gamma))$$

for some modulus of continuity Φ.

Deficit bounds without distances

Carlen (1991)

$$\delta_c(f) = \frac{1}{2\pi}\int |\nabla f|^2\, dm - \int |f|^2 \log |f|^2\, dm \geq \int |\mathcal{W} f|^2 \log |\mathcal{W} f|^2\, dm \geq 0.$$

Note that $\delta(f) = \delta(u_f)$ with $u_f(x) = (f(2\sqrt{\pi}x))^{\frac{1}{2}}$.

Ledoux (1992)

$$\delta(f) = \int_0^\infty e^{-4t} \int \frac{1}{|P_t f|^3} \sum_{i,j=1}^{n} (P_t f_i P_t f_j - P_t f P_t(f_i f_j))^2\, d\gamma dt \geq 0.$$

Cordero-Erausquin (2002)

$$\delta(f) \geq \int |\nabla(\log f) + \nabla\theta|^2 f d\gamma + \int (\Delta_A \theta - \log\det(I + \operatorname{Hess}\theta))\, f d\mu.$$

W_2-stability via optimal transportation

The p-th Wasserstein distance between μ and ν is

$$W_p(\mu,\nu) = \inf_{\pi} \left(\iint_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^p \, d\pi(x,y) \right)^{\frac{1}{p}}$$

where the infimum is taken over all couplings of μ and ν.

Theorem (Indrei, Marcon 2014)

If $f d\gamma$ is a centered probability measure such that

$$-1 + \varepsilon \leq D^2(\log(1/f)) \leq M$$

for some $M, \varepsilon > 0$, then we have $\delta(f) \geq C_{\varepsilon,M} W_2^2(f d\gamma, d\gamma)$.

Remark: $D^2(\log(1/f)) \geq -1 + \varepsilon$ implies $f d\gamma$ is log concave.

Idea: Optimal transport deficit bound

$$\delta(f) \geq \int (\Delta_A \theta - \log\det(I + \operatorname{Hess}\theta)) \, f d\mu$$

with Caffarelli's contraction theorem.

W_2 and L^1 stability via Semigroup method

Theorem (Fathi, Indrei, Ledoux 2016)

Suppose that $fd\gamma$ is centered and satisfies a $(2,2)$-Poincare inequality with $\lambda > 0$, i.e.,

$$\lambda \int |g|^2 f d\gamma \leq \int |\nabla g|^2 f d\gamma$$

for every smooth function g. Then

$$\frac{c(\lambda)}{2} \mathrm{I}(f) \geq \mathrm{H}(f)$$

for some $c(\lambda) = \frac{1-\lambda+\lambda\log\lambda}{(1-\lambda)^2} < 1$. (Note that $\delta(f) = \frac{1}{2}\mathrm{I}(f) - \mathrm{H}(f)$.)
As a consequence,

$$\delta(f) \geq C_1(\lambda) W_2^2(fd\gamma, d\gamma)$$
$$\delta(f) \geq C_2(\lambda) \|f-1\|_{L^1(d\gamma)}.$$

Remark

(1) In $n = 1$, the class of probability measures satisfying Poincare inequality is fully characterized.

(2) For higher dimension, there is a general sufficient condition for Poincare inequality. If f is log-concave, then $fd\gamma$ satisfies $(2,2)$-Poincare inequality.

L^1 stability via Fourier analysis

Recall: Carlen (1991)

$$\frac{1}{2\pi}\int |\nabla f|^2\,dm - \int |f|^2 \log|f|^2\,dm \geq \int |\mathcal{W}f|^2 \log|\mathcal{W}f|^2\,dm$$
$$\geq \frac{1}{2}\left(\int ||\mathcal{W}f| - 1|^2\,dm\right)^2.$$

Theorem (Feo–Indrei–Posteraro–Roberto 2017)

If $f d\gamma$ is a probability measure satisfying

$$\mathcal{F}(e^{-\pi|x|^2} f(2\sqrt{\pi}x)) \geq 0,$$

then

$$\delta(f) \geq \frac{1}{2}\|f-1\|_2^4.$$

W_2-stability with moment bounds

Theorem (Bobkov–Gozlan–Roberto–Samson 2014)

If $f d\gamma$ is centered and $\int |x|^2 f d\gamma \leq n$, then

$$\delta(f) \geq \frac{C}{n} W_2^4(f d\gamma, d\gamma).$$

Remark

If f satisfies $(2,2)$ Poincare with λ, then

$$\int |x|^2 f \, d\gamma \leq \frac{1}{\lambda} \int |\nabla x|^2 f \, d\gamma = \frac{n}{\lambda}.$$

Deficit bound via Scaling asymmtry

Euclidean LSI (equivalent to Gaussian LSI) states

$$\int |\nabla h|^2 \, dx \geq \frac{1}{2} \int h^2 \log h^2 \, dx + \frac{n}{4} \log(2\pi e^2).$$

If $g(x) = t^{-\frac{n}{2}} h(x/t)$, then

$$\int |\nabla h|^2 \, dx = t^2 \int |\nabla g|^2 \, dx,$$

$$\int h^2 \log h^2 \, dx = \int g^2 \log g^2 \, dx + n \log t.$$

Theorem (Dolbeault, Toscani 2016)

$$\delta(f) \geq \frac{1}{4} \left((n - m_2(f)) + 2\mathrm{H}(f) \right)^2$$

where $m_2(f) = \int |x|^2 f \, d\gamma$ is the second moment.

In particular, if $m_2(f) \leq n$ then L^1-stability follows from Pinsker's inequality.

Stability results under Moment assumptions

with respect to L^1 and W_1

For $M > 0$, let $\mathcal{P}_2^M(\mathbb{R}^n)$ be the space of probability measures μ such that

$$\int_{\mathbb{R}^n} |x|^2 \, d\mu \leq M.$$

Theorem (Indrei–K. 2018)

Let $f d\gamma, f_k d\gamma \in \mathcal{P}_2^M(\mathbb{R}^n)$ be centered.

(1) For $n = 1$,

$$\delta(f) \geq C_M \|f - 1\|_{L^1(d\gamma)}^4.$$

(2) For $n \geq 2$,

$$\delta(f_k) \to 0 \quad \Longrightarrow \quad \|f_k - 1\|_{L^1(d\gamma)} \to 0.$$

(3) For $n \geq 1$,

$$\delta(f) \geq C_{n,M} \min\{W_1(f d\gamma, d\gamma), W_1^4(f d\gamma, d\gamma)\}.$$

Stability results under Moment assumptions

Key ideas

Using HWI inequality

$$\mathrm{H}(f) \leq W_2(fd\gamma, d\gamma)\sqrt{\mathrm{I}(f)} - \frac{1}{2}W_2^2(fd\gamma, d\gamma),$$

we derive

$$\mathrm{H}(f), \mathrm{I}(f) \leq C_1\delta(f) + C_2(n, M)$$

when $fd\gamma \in \mathcal{P}_2^M$. Then,

(1) By the compactness argument (Rellich-Kondrashov, since $\mathrm{I}(f)$ is bounded),

$$\delta(f_k) \to 0 \quad \Longrightarrow \quad \|f_k - 1\|_{L^1(d\gamma)} \to 0.$$

(2) By W_1-stability of Talagrand's transportation inequality by Cordero-Erausquin,

$$\delta(f) \geq C_{n,M} \min\{W_1(fd\gamma, d\gamma), W_1^4(fd\gamma, d\gamma)\}.$$

Stability results under density bounds assumptions

Using the deficit bound from optimal transport

For $g \in L^1(d\gamma)$ and $\alpha > 0$, we define

$$\mathcal{B}(\alpha) = \{f d\gamma \in \mathcal{P} : f(x) \geq \alpha \text{ a.e. } x\},$$
$$\mathcal{B}(\alpha, g) = \{f d\gamma \in \mathcal{P} : \alpha \leq f(x) \leq g(x) \text{ a.e. } x\}$$

where $\mathcal{P}$ is the space of all probability measures.

Theorem (Indrei–K. 2018)

Let $\alpha \in (0, 1]$ and $f d\gamma \in \mathcal{B}(\alpha)$ be centered. Then there exists a linear function $L_f = a_f \cdot x + b_f$ such that

$$\delta(f) \geq C(\alpha, n) \|\log f - L_f\|^2_{L^1(d\gamma)},$$

where $a_f \in \mathbb{R}^n$, $b_f \in \mathbb{R}$, and $|a_f| + |b_f| \leq c(n, \alpha)$.

If $\{f_k d\gamma\} \subset \mathcal{B}(\alpha, g)$ for some $g \in L^1(d\gamma)$ and $\delta(f_k) \to 0$, then $f_k \to 1$ in $L^1(d\gamma)$.

Stability combining Fourier analysis and Optimal transport

Theorem (Indrei–K. 2018)

If $f \in L^2(dm)$ with $\|f\|_{L^2(dm)} = 1$, then

$$\sqrt{2n}\delta_c^{\frac{1}{2}}(f) + \delta_c(f) \geq 2\pi \int |x|^2 dm - 2\pi \int |x|^2 |f|^2 dm + \frac{1}{2\pi} \int |\nabla f|^2 dm.$$

This implies weak L^2–stability in $\mathcal{P}_2^M(\mathbb{R}^n)$ for δ_c (equivalently, weak L^1–stability for δ).

Key step

We consider the Brenier map between dm and $|\mathcal{W} f|^2 \, dm$.

Relax assumptions and stronger distances

Question: Can we relax the assumptions on probability mesures or get stronger distance bounds?

W_2-stability

(1) $(2,2)$-Poincare inequality with $\lambda > 0$ [Fathi et al 2016].

(2) the second moment $\int |x|^2 f \, d\gamma$ is bounded by n [Bobkov et al 2014].

L^p-stability in $\mathcal{P}_2^M$

(1) In $n = 1$, Quantitative stability in L^1.

(2) In $n \geq 2$, Non-quantitative stability in L^1.

Instability results

Theorem (K. 2018)

Let $M > n$ and $p > 1$. There exists a sequence of centered probability measures $f_k d\gamma \in \mathcal{P}_2^M(\mathbb{R}^n)$ such that $\lim_{k\to\infty} \delta(f_k) = 0$,

$$\liminf_{k\to\infty} W_2(f_k d\gamma, d\gamma) \geq C_1,$$

and

$$\liminf_{k\to\infty} \|f_k - 1\|_{L^p(d\gamma)} \geq C_2,$$

for some $C_1, C_2 > 0$.

Instability results

Recall that if $fd\gamma \in \mathcal{P}_2^M(\mathbb{R}^n)$, then

$$\delta(f) \geq C_{n,M} \min\{W_1(fd\gamma, d\gamma), W_1^4(fd\gamma, d\gamma)\}.$$

Let $\mathcal{P}_2(\mathbb{R}^n) = \cup_{M>0}\mathcal{P}_2^M(\mathbb{R}^n)$.

Theorem (K. 2018)

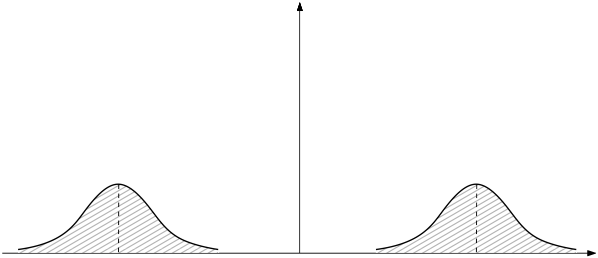
There exists a sequence of centered probability measures $f_k d\gamma \in \mathcal{P}_2(\mathbb{R})$ such that $\lim_{k\to\infty} \delta(f_k) = 0$ and $\lim_{k\to\infty} W_1(f_k d\gamma, d\gamma) = \infty$.

Construction of Examples

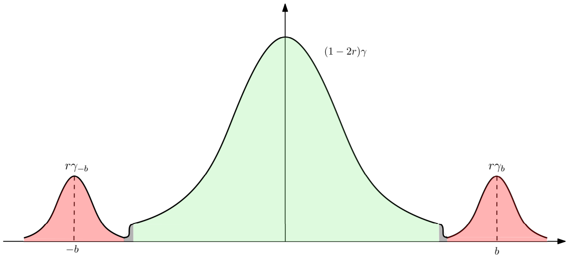
For any c_1, c_2, our goal is to construct a sequence of centered probability measures $f_k d\gamma$ such that $\delta(f_k) \to 0$ but $W_1, W_2, L^p \not\to 0$. To this end, it suffices to control the relative entropy $\mathrm{H}(f)$ and the p-th moment $m_p(f)$ because

$$\frac{1}{2}\mathrm{H}(f) \leq \frac{2}{p-1}\|f-1\|_1^p + 2\|f-1\|_p,$$

$$m_p(\mu) - m_p(\nu) \geq C_1 > 0 \qquad \Longrightarrow \qquad W_p(\mu,\nu) \geq C_2 > 0.$$



Construction of Examples



Controlling the barycenters and the mass of two bumps, we obtain small deficit, large distance, in a class of centered probability measures with bounded second moments.

Remark: Similar examples (mixture of Gaussian) were studied by [Eldan, Lehec, Shenfeld (2020)].

Thank you!