Hypercontractivity and Logarithmic Sobolev Inequalities and their Applications

Salman Beigi

School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

January 22, 2021

Any stochastic map T is a contraction under L_p-norms, namely $\|Tf\|_p \leq \|f\|_p$ for any $1 \leq p \leq \infty$ and arbitrary f. This is a simple consequence of the convexity of $x \mapsto x^p$. Some stochastic maps satisfy stronger inequalities of the form

$$\|Tf\|_p \leq \|f\|_q, \qquad \forall f, \tag{1}$$

for some $1 \leq q < p \leq \infty$. This is a stronger inequality since $q \mapsto \|f\|_q$ is a non-decreasing function. Thus (1) is stronger than the contraction of T, so is called a *hypercontractivity inequality*. Proving hypercontractivity inequalities is usually a challenge, yet when T belongs to a continuous semigroup of stochastic maps, hypercontractivity inequalities can be derived from another type of inequalities called *logarithmic Sobolev inequalities*.

Hypercontractivity and logarithmic Sobolev inequalities have found several applications and connections to other areas of mathematics such as concentration of measure inequalities, transportation cost inequalities, isoperimetric inequalities, bounding the mixing times and analysis of boolean functions. In this manuscript we introduce the notations of hypercontractivity and logarithmic Sobolev inequalities and survey some of their applications and connections to other fields.

Contents

1 Markov Semigroups

Let (Ω, Σ, π) be a probability space. To avoid some technicalities for now we assume that Ω is a finite set and π has full support. Let $L^2(\pi) = \{f : \Omega \to \mathbb{R}\}$ be the space of real functions on Ω equipped with the inner product

$$\langle f, g\rangle_\pi = \mathbb{E}[fg], \tag{2}$$

where the expectation is with respect to π. This inner product induces the norm

$$\|f\|_2 = (\mathbb{E}f^2)^{1/2}.$$

By abuse of notation here a real number $c \in \mathbb{R}$ is also considered as an element of $L^2(\pi)$ (the constant function). In particular 1 is the constant 1 function.

The variance of a function f is equal to

$$\mathrm{Var} f = \mathbb{E}[f^2] - \mathbb{E}[f]^2.$$

A collection $\{T_t : t \geq 0\}$ of maps $T_t : L^2(\Omega) \to L^2(\Omega)$ is called a *Markov semigroup* if it satisfies:

(a) Semigroup: $T_0 = I$ (the identity map) and $\forall t, s \geq 0$ we have

$$T_s T_t = T_{s+t},$$

(b) Continuous: $t \mapsto T_t$ is continuous in the following sense: for all functions f and all $x \in \Omega$ we have

$$\lim_{t\to 0^+} T_t f(x) = f(x),$$

(c) Stochastic: T_t for all $t \geq 0$ is a stochastic map, i.e.,

$$\begin{array}{lll} \text{(normalization)} & T_t 1 = 1 & \\ \text{(positivity)} & T_t f \geq 0 & \forall f \geq 0, \end{array}$$

where $f \geq 0$ means that $f(x) \geq 0$ for all $x \in \Omega$.

Observe that by the normalization condition 1 is a common eigenvalues of T_t's. Using the positivity condition, it can be shown that 1 is indeed the largest eigenvalues of T_t's. Thus these operators are uniformly bounded.

The way we define a stochastic map T_t, it acts on the space of functions. Then its *transpose* denoted by T^* acts on the space of probability distributions. In other words, for any probability distribution μ, which can be thought of as a row vector,

$$\tau = \mu T_t,$$

is again a probability distribution. Indeed, $\tau = \mu T_t$ is coordinate-wise non-negative since by the positivity condition $\tau f = \mu(T_t f) \geq 0$ for all $f \geq 0$, and is normalized because $\tau 1 = \mu(T_t 1) = \mu 1 = 1$.

Lindblad operator: The Lindblad operator (or generator) associate to a Markov semigroup $\{T_t : t \geq 0\}$ is defined by

$$\mathcal{L} := -\lim_{t\to 0^+} \frac{1}{t}(T_t - I).$$

Here we prove that this limit exists. By the continuity of $t \mapsto T_t$ for sufficiently small $\epsilon > 0$ we have

$$\Big\| I - \frac{1}{\epsilon}\int_0^\epsilon T_s \mathrm{d}s \Big\| < 1,$$

where $\|\cdot\|$ denotes the operator norm. Then $\epsilon^{-1}\int_0^\epsilon T_s\mathrm{d}s$, and then $\int_0^\epsilon T_s\mathrm{d}s$ are invertible. Now for $0 < t < \epsilon$ we have

$$\begin{aligned}
\frac{1}{t}(T_t - I)\int_0^\epsilon T_s\mathrm{d}s &= \frac{1}{t}\Big(\int_0^\epsilon T_{s+t}\mathrm{d}s - \int_0^\epsilon T_s\mathrm{d}s\Big)\\
&= \frac{1}{t}\Big(\int_t^{t+\epsilon} T_s\mathrm{d}s - \int_0^\epsilon T_s\mathrm{d}s\Big)\\
&= \frac{1}{t}\Big(\int_\epsilon^{t+\epsilon} T_s\mathrm{d}s - \int_0^t T_s\mathrm{d}s\Big).
\end{aligned}$$

Multiplying both sides from right by the inverse of $\int_0^\epsilon T_s\mathrm{d}s$ we find that

$$\frac{1}{t}(T_t - I) = \frac{1}{t}\Big(\int_\epsilon^{t+\epsilon} T_s\mathrm{d}s - \int_0^t T_s\mathrm{d}s\Big)\cdot\Big(\int_0^\epsilon T_s\mathrm{d}s\Big)^{-1}.$$

Then taking the limit by continuity we get

$$\lim_{t\to 0^+}\frac{1}{t}(T_t - I) = (T_\epsilon - I)\Big(\int_0^\epsilon T_s\mathrm{d}s\Big)^{-1}.$$

As a result the limit exists and we have

$$\mathcal{L} = -(T_\epsilon - I)\Big(\int_0^\epsilon T_s\mathrm{d}s\Big)^{-1}.$$

□

Indeed using the semigroup property of T_t's we have

$$\frac{\mathrm{d}}{\mathrm{d}t}T_t = -\mathcal{L}T_t = -T_t\mathcal{L}.$$

and then,

$$T_t = e^{-t\mathcal{L}}.$$

From our assumption that $T_t 1 = 1$ we immediately find that

$$\mathcal{L}1 = 0.$$

Reversibility: Throughout this manuscript we assume that $\mathcal{L}$ is self-adjoint as an operator acting on $L^2(\pi)$. That is, for all f, g we have

$$\langle f, \mathcal{L}g\rangle_\pi = \langle \mathcal{L}f, g\rangle_\pi,$$

where the inner product is defined in (2). The fact hat $\mathcal{L}$ is self-adjoint essentially means that the semigroup is *reversible* (or satisfies the *detailed balance* condition).

If the Markov semigroup is reversible, then π is a *stationary distribution* of T_t for all $t \geq 0$. To see this, note that if $\mathcal{L}$ is self-adjoint, then $T_t = e^{-t\mathcal{L}}$ is self-adjoint, i.e.,

$$\langle f, T_t g\rangle_\pi = \langle T_t f, g\rangle_\pi.$$

Now letting f, g be the characteristic functions of the sets $\{x\}, \{y\}$ for some $x, y \in \Omega$, we find that

$$\pi(x)T_t(x, y) = \pi(y)T_t(y, x). \tag{3}$$

Now we have

$$\pi T_t(y) = \sum_x \pi(x)T_t(x, y) = \sum_x \pi(y)T_t(y, x) = \pi(y),$$

where in the last equation we use $T_t 1 = 1$. Then $\pi T_t = \pi$ and π is the stationary distribution of T_t.

Since $\mathcal{L}$ and T_t are self-adjoint, T_t does not change the average with respect to π:

$$\mathbb{E}[T_t f] = \langle 1, T_t f\rangle_\pi = \langle T_t 1, f\rangle_\pi = \langle 1, f\rangle_\pi = \mathbb{E}f.$$

We similarly have $\mathbb{E}[\mathcal{L}f] = 0$ for all f.

Assume that μ is another probability distribution and let $f = \mu/\pi$. Also let $\tau = \mu T_t$ and $g = \tau/\pi$. Then we have

$$\begin{aligned} g(x) &= \frac{\tau(x)}{\pi(x)} \\ &= \sum_y \frac{\mu(y)T_t(y, x)}{\pi(x)} \\ &= \sum_y \frac{\mu(y)T_t(x, y)}{\pi(y)} \\ &= T_t f(x), \end{aligned}$$

where in the third line we use the reversibility condition (3). In summary, for reversible Markov semigroups we have

$$\frac{\mu T_t}{\pi} = T_t\Big(\frac{\mu}{\pi}\Big). \tag{4}$$

Dirichlet form: The Dirichlet form associate with the semigroup is defined by

$$\mathcal{E}(f,g) := \langle f, \mathcal{L}g\rangle_\pi = \mathbb{E}[f\mathcal{L}g] = -\frac{\mathrm{d}}{\mathrm{d}t}\langle f, T_t g\rangle_\pi\Big|_{t=0}.$$

We claim that the Dirichlet form is positive, namely $\mathcal{E}(f,f) \geq 0$ for all f. Equivalently, as $\mathcal{L}$ is self-adjoint, $\mathcal{L}$ is positive semidefinite. To prove this, let $r < 0$, and define

$$\hat{T}_t = e^{t(rI-\mathcal{L})} = e^{rt}T_t.$$

We then have

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{T}_t = (rI - \mathcal{L})\hat{T}_t,$$

and

$$(rI - \mathcal{L})\int_0^t \hat{T}_s \mathrm{d}s = \hat{T}_t - I.$$

On the other hand since $r < 0$ and T_t is bounded, $\lim_{t\to\infty} \hat{T}_t = 0$. Therefore,

$$(rI - \mathcal{L})\int_0^\infty \hat{T}_s \mathrm{d}s = -I.$$

This means that $rI - \mathcal{L}$ in invertible, and then r is not an eigenvalue of $\mathcal{L}$. That is, $\mathcal{L}$ is self-adjoint and all of whose eigenvalues are non-negative, so it is positive semidefinite.

$\square$

Following similar ideas, the following theorem of Hille and Yosida can be proven.

Theorem 1.1. *(Hille-Yosida theorem) For any self-adjoint operator $\mathcal{L}$ that is the generator of a Markov semigroup we have*

$$\|(aI + \mathcal{L})^{-1}\| \leq \frac{1}{a},$$

for any $a > 0$. Equivalently for any vector f and $a > 0$ we have

$$a\|f\|_2 \leq \|(aI + \mathcal{L})f\|_2.$$

The above theorem holds in the infinite dimensional case as well. In the finite dimensional case that we consider here, however, there is a simpler characterization of Lindblad operators. We claim that $\mathcal{L}$ satisfies

(i) $\mathcal{L}1 = 0$,

(ii) off-diagonal entries of $\mathcal{L}$ are at most zero.

if and only if $\{T_t = e^{-t\mathcal{L}} : t \geq 0\}$ form a Markov semigroup. The continuity and semigroup properties obviously hold from the definition $T_t = e^{-t\mathcal{L}}$. The normalization condition is also equivalent to (i). It remains to prove that the positivity condition is equivalent to (ii). For one direction we note that since the off-diagonal entries of $T_0 = I$ are all zero, and the off-diagonal entries of T_t are non-negative, $\mathcal{L}$, as the minus derivative of T_t at zero, must satisfy (ii). For the other direction, assuming (ii), there is some positive $a > 0$ such that all entries of $aI - \mathcal{L}$ are non-negative. Then $T_t = e^{-t\mathcal{L}} = e^{-at}e^{t(aI-\mathcal{L})}$ is a positive constant times the exponential of a matrix with non-negative entries. This means that T_t has non-negative entries.

Spectral gap: We saw that $\mathcal{L}$ is positive semidefinite and has a zero eigenvalue since $\mathcal{L}1 = 0$. Hereafter we assume that 1 is the only 0-eigenvector of $\mathcal{L}$, i.e., the zero eigenvalue is non-degenerate. This condition is called the *primitivity* condition.

Having a primitive semigroup, its *spectral gap* is the smallest non-zero eigenvalue of $\mathcal{L}$ which we denote by λ. Since in this case the function 1 is the sole 0-eigenvalue of $\mathcal{L}$ and the space of functions orthogonal to 1 is $\{f : \mathbb{E}f = 0\}$ we have

$$\lambda = \inf_{\mathbb{E}[f]=0} \frac{\mathcal{E}(f, f)}{\mathbb{E}[f^2]}. \tag{5}$$

As an exercise it can also be shown that

$$\lambda = \inf_{f\neq 0} \frac{\mathcal{E}(f, f)}{\mathrm{Var} f}.$$

Proposition 1.2. *For any $t \geq 0$ and any function f we have*

$$\|T_t f - \mathbb{E}f\|_2 \leq e^{-\lambda t}\|f - \mathbb{E}f\|_2.$$

Proof. By the definition of the spectral gap for every function f we have

$$\mathrm{Var} f = \|f - \mathbb{E}f\|_2^2 \leq \frac{1}{\lambda}\mathcal{E}(f, f).$$

Such an inequality is called a *Poincare inequality.* Now we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\|T_t f - \mathbb{E}f\|_2^2 &= \frac{\mathrm{d}}{\mathrm{d}t}\langle T_t f - \mathbb{E}f, T_t f - \mathbb{E}f\rangle \\
&= -2\mathcal{E}(T_t f, T_t f) \\
&\leq -2\lambda\|T_t f - \mathbb{E}f\|_2^2.
\end{aligned}$$

Taking the integration this gives the desired inequality. □

This proposition says that, assuming that $\mathcal{L}$ is primitive and $\lambda > 0$, for any f, as $t \to \infty$, $t \mapsto T_t f$ tends to a constant function equal to its expectation. Equivalently, by (4), for any probability distribution μ, μT_t converges to the stationary distribution π. Moreover, this convergence is exponentially fast.

To summarize, in the rest of this manuscript we assume that $\mathcal{L}$ is a Lindblad generator of a Markov semigroup with elements

$$T_t = e^{-t\mathcal{L}}.$$

We assume that $\mathcal{L}1 = 0$ (equivalently, T_t satisfies the normalization condition) and $T_t f \geq 0$ for all $f \geq 0$ so that $\{T_t : t \geq 0\}$ is a valid Markov semigroup. We further assume that $\mathcal{L}$ satisfies the reversibility or detailed balance condition, namely $\mathcal{L}$ is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$. We showed that in this case $\mathcal{L}$ is positive semidefinite (although has a 0-eigenvalue) and defined associated Dirichlet form $\mathcal{E}(\cdot, \cdot)$. We further assume that $\mathcal{L}$ is primitive, i.e., the 0-eigenvalue of $\mathcal{L}$ is non-degenerate and denote the spectral gap of $\mathcal{L}$ by λ.

Let us finish this section by giving a simple yet important example. Let $\mathcal{L} = I - \mathbb{E}$ where the expectation $\mathbb{E}$ is with respect to some fixed distribution π, i.e.,

$$\mathcal{L}f = f - \mathbb{E}f.$$

Then we have

$$T_t f = e^{-t(I-\mathbb{E})}f = e^{-t}e^{t\mathbb{E}}f = e^{-t}(I + (e^t - 1)\mathbb{E})f = e^{-t}f + (1 - e^{-t})\mathbb{E}f,$$

where in computing $e^{t\mathbb{E}}$ we use the fact that $\mathbb{E}^2 = \mathbb{E}$ is a projection. It is easily verified that $\{T_t : t \geq 0\}$ is a Markov semigroup. Moreover, we have

$$\langle f, \mathcal{L}g\rangle_\pi = \langle f, g - \mathbb{E}g\rangle_\pi = \mathbb{E}[f(g - \mathbb{E}g)] = \mathbb{E}[fg] - \mathbb{E}[f]\mathbb{E}[g] = \langle \mathcal{L}f, g\rangle_\pi.$$

Thus $\mathcal{L}$ is reversible. Observe that

$$\mathcal{E}(f, f) = \langle f, \mathcal{L}f\rangle_\pi = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \mathrm{Var}f \geq 0,$$

so the Dirichlet form $\mathcal{E}(\cdot, \cdot)$ is positive. Finally, note that $\mathbb{E}$ is a rank-one projection. Thus $\mathcal{L} = I - \mathbb{E}$ has a single 0-eigenvalue and all of whose other eigenvalues are 1. Therefore, $\lambda = 1$ and $\mathcal{L}$ is primitive.

2 Hypercontractivity inequalities

For any $p \geq 1$ we define the p-norm by

$$\|f\|_p := (\mathbb{E}|f|^p)^{1/p},$$

which generalized the definition of 2-norm given in the previous section. We emphasis that here the expectation $\mathbb{E}$ is with respect to the fixed probability distribution π. We also define $\|f\|_\infty$ in limit of $p \to \infty$ of $\|f\|_p$, i.e., we let $\|f\|_\infty = \max_x |f(x)|$. By Minkowski's inequality $\|\cdot\|_p$, for $p \geq 1$, satisfies triangle's inequality and indeed is a norm.

For any p define $\hat{p}$ by

$$\frac{1}{p} + \frac{1}{\hat{p}} = 1. \tag{6}$$

$\hat{p} = p/(p-1)$ is called the *Hölder conjugate* of p. Then *Hölder's duality* states that for $p \geq 1$ (and then $\hat{p} \geq 1$) we have

$$\|f\|_p = \max_{g \neq 0} \frac{\langle f, g\rangle_\pi}{\|g\|_{\hat{p}}},$$

which in particular says that

$$|\langle f, g\rangle_\pi| \leq \|f\|_p \cdot \|g\|_{\hat{p}}.$$

Hölder's inequality for $p = 2$ is usually called the Cauchy-Schwarz inequality. In general, Hölder's inequality can be proven using Young's inequality.

Entropy: A crucial property of p-norms is their connection to the *entropy* function. For any function $f \geq 0$ we define its entropy by

$$\mathrm{Ent}_\pi(f) := \mathbb{E}[f \log f] - \mathbb{E}f \log \mathbb{E}f.$$

Using the convexity of $s \mapsto s \log s$ we have $\mathrm{Ent}(f) \geq 0$ for all $f \geq 0$.

The entropy function as defined above is related to the KL-divergence (relative entropy) as follows. For two distributions μ, π their KL-divergence is defined by

$$D(\mu \| \pi) = \sum_x \mu(x)\big(\log \mu(x) - \pi(x)\big). \tag{7}$$

Now observe that for the function $f = \mu/\pi$ we have

$$\mathrm{Ent}_\pi(f) = D(\mu \| \pi).$$

The relevance of the entropy function to hypercontractivity inequalities is due to the following proposition which can be proven by a simple computation.

Proposition 2.1. *For any $f \geq 0$ we have*

$$\frac{\mathrm{d}}{\mathrm{d}p}\|f\|_p = \frac{1}{p^2}\|f\|_p^{1-p}\ \mathrm{Ent}_\pi(f^p).$$

Note that as a corollary of this proposition and the non-negativity of the entropy function we find that $p \mapsto \|f\|_p$ is non-decreasing:

$$\frac{\mathrm{d}}{\mathrm{d}p}\|f\|_p \geq 0.$$

Operator norm: For $p, q \geq 1$, the $q \to p$ norm of T_t is defined by

$$\|T_t\|_{q\to p} := \sup_{f \neq 0} \frac{\|T_t f\|_p}{\|f\|_q}.$$

In other words, $\|T_t\|_{q\to p}$ is the smallest number M such that

$$\|T_t f\|_p \leq M\|f\|_q, \qquad \forall f.$$

Note that since the entries of T_t are non-negative, the above maximum is achieved at some $f \geq 0$. For such f using the convexity of $s \mapsto s^q$ one can easily verify that $\|T_t f\|_q^q \leq \|f\|_q^q$. Therefore,

$$\|T_t\|_{q\to q} \leq 1 \qquad \forall q \geq 1.$$

That is, T_t is a *contraction* under any q-norm for $q \geq 1$.

We say that T_t is *hypercontractive* if $\|T_t\|_{q\to p} \leq 1$, or equivalently

$$\|T_t f\|_p \leq \|f\|_q,$$

for some $1 \le q < p$. Note that since $p \mapsto \|f\|_p$ is non-decreasing as mentioned above, such an inequality is stronger than the contractivity inequality $\|T_t\|_{q\to q} \le 1$. This is why such an inequality is called a hypercontractivity inequality.

Hypercontractivity inequalities are usually challenging to prove. However, since here T_t belongs to a Markov semigroup we may use other tools for proving such inequalities.

Definition 2.2. For an arbitrary q, we say that the Markov semigroup $\{T_t : t \ge 0\}$ satisfies the q-log-Sobolev inequality with constant $c > 0$ if for all $f > 0$ we have

$$c \operatorname{Ent}_\pi(f^q) \le \frac{q^2}{4(q-1)} \mathcal{E}(f^{q-1}, f),$$

where as before $\mathcal{E}(f, g) = \langle f, \mathcal{L}g\rangle_\pi$. The log-Sobolev inequality for $q = 1$ is defined (as the limit of $q \to 1^+$) by

$$c \operatorname{Ent}_\pi(f) \le \frac{1}{4} \mathcal{E}(f, \log f).$$

We denote the best constant c satisfying the above log-Sobolev inequality by α_q.

In a log-Sobolev inequality, the entropy function appears in the left hand side, whose definition involves the logarithm function. This is why it is called a logarithmic Sobolev inequality. While the left hand side of such an inequality does not depend on the semigroup, the Dirichlet form appears in the right hand side and depends on the generator of the Markov semigroup.

Letting $f = g^{1/q}$ and using the definition of Hölder's conjugate $\hat{q} = q/(q-1)$ the q-log-Sobolev inequality is equivalent to

$$c \operatorname{Ent}_\pi(g) \le \frac{q\hat{q}}{4} \mathcal{E}(g^{1/\hat{q}}, g^{1/q}), \qquad \forall g > 0. \tag{8}$$

Also for the q-log-Sobolev constant we have

$$\alpha_q = \inf_g \frac{q\hat{q}\mathcal{E}(g^{1/\hat{q}}, g^{1/q})}{4 \operatorname{Ent}_\pi(g)}. \tag{9}$$

From this expression it is clear that $\alpha_q = \alpha_{\hat{q}}$.

Theorem 2.3. (i) *For any $1 \le q \le p \le 2$ we have*

$$q\hat{q}\mathcal{E}(g^{1/\hat{q}}, g^{1/q}) \ge p\hat{p}\mathcal{E}(g^{1/\hat{p}}, g^{1/p}), \qquad \forall g \ge 0.$$

(ii) *The function $q \mapsto \alpha_q$ is non-increasing on $[1, 2]$. In particular α_2 is the smallest log-Sobolev constant.*

Proof. (ii) is a simple consequence of (i) and the definition of log-Sobolev constants. The inequality in (i) is called the *Stroock-Varopoulos inequality*, whose proof can be found e.g. in [12, Theorem 2.1]. Here we given a different proof. For any $t \ge 0$ and $f \ge 0$ define

$$h_t(s) = \langle f^{2-s}, T_t f^s\rangle_\pi.$$

Observe that $h_t(2-s) = h_t(s)$ so that h_t is symmetric about $s = 1$. Therefore all the the odd-order derivatives of h_t at $s = 1$ vanish and for the Taylor expansion of h_t at $s = 1$ we have

$$h_t(s) = h_t(1) + \sum_{j=1}^{\infty} \frac{c_j}{(2j)!}(s-1)^{2j},$$

with

$$c_j = \frac{\mathrm{d}^{2j}}{\mathrm{d}s^{2j}} h_t(s)\Big|_{s=1}.$$

By a simple computation we have

$$h_t(s) = \sum_{x,y} \pi(x) T_t(x,y) f(x)^2 e^{s \log(f(y)/f(x))}.$$

That is $h_t(s)$ is a summation of exponential functions with positive coefficients. From this we obtain $c_j \geq 0$ for all j.

Now define

$$\psi_t(s) = \frac{h_t(s) - h_t(0)}{(s-1)^2 - s} = \sum_{j=1}^{\infty} \frac{c_j}{(2j)!}\Bigg(\sum_{i=1}^{j-1} (s-1)^{2i}\Bigg).$$

From this expression and non-negativity of c_j's we find that $\psi_t(s)$ is non-decreasing on $[1, \infty)$. Therefore, $\lim_{t\to 0^+} \psi_t(s)/t$ is non-decreasing on the same interval. On the other hand, using $h_t(0) = \mathbb{E}[f^2] = h_0(s)$ we compute

$$\begin{aligned}
\lim_{t\to 0^+} \frac{\psi_t(s)}{t} &= \frac{1}{(s-1)^2 - 1} \lim_{t\to 0^+} \frac{h_t(s) - h_t(0)}{t} \\
&= \frac{1}{(s-1)^2 - 1} \lim_{t\to 0^+} \frac{h_t(s) - h_0(s)}{t} \\
&= \frac{1}{(s-1)^2 - 1} \frac{\partial}{\partial t} h_t(s)\Big|_{t=0} \\
&= -\frac{1}{(s-1)^2 - 1} \big\langle f^{2-s}, \mathcal{L} f^s \big\rangle_\pi.
\end{aligned}$$

Therefore

$$s \mapsto -\frac{1}{(s-1)^2 - 1} \big\langle f^{2-s}, \mathcal{L} f^s \big\rangle_\pi,$$

is non-decreasing on $[1, +\infty)$. Now the desired result follows once we identify $2/s$ with p (and $2/(2-s)$ with $\hat{p}$, its Hölder conjugate) and $f = \sqrt{g}$.

□

We can now state the main result of this section.

Theorem 2.4. *Let $\mathcal{L}$ be the generator of a reversible and primitive Markov semigroup.*

(i) *We have*

$$\|T_t\|_{q\to p} \leq 1, \qquad \forall p, q > 1, \quad \frac{p-1}{q-1} \leq e^{4\alpha_2 t}.$$

(ii) *Conversely if for some $q > 1$ we have*

$$\|T_t\|_{q\to p} \le 1, \qquad \forall p > 1, \quad \frac{p-1}{q-1} \le e^{4ct},$$

then $\alpha_q \ge c$.

Proof. (i) Let $q > 1$ be arbitrary and define $t(p) = \frac{1}{4\alpha_2}\log\frac{p-1}{q-1}$. For some $f \ge 0$ define

$$\psi(p) = \|f\|_q - \|T_{t(p)}f\|_p.$$

By a straightforward computation (using Proposition 2.1) we have

$$\psi'(p) = -\frac{1}{p^2}\mathbb{E}[g_p]^{\frac{1-p}{p}}\Big(\mathrm{Ent}_\pi(g_p) - \frac{p^2}{4\alpha_2(p-1)}\langle g_p^{1/\hat{p}}, g_p^{1/p}\rangle_\pi\Big),$$

where $g_p = \big(T_{t(p)}f\big)^p$. Now since by Theorem 2.3 we have $\alpha_q \ge \alpha_2$. Therefore, $\psi'(p) \ge 0$. Therefore, for any $p \ge q$ we have

$$\psi(p) \ge \psi(q) = 0.$$

This gives the desired hypercontractivity inequality.

(ii) The proof is similar to that of part (i). Let $t(p) = \frac{1}{4\alpha_2}\log\frac{p-1}{q-1}$ and

$$\psi(p) = \|f\|_q - \|T_{t(p)}f\|_p.$$

Then by assumption we have $\psi(p) \ge 0$ for all $p \ge q$. We also have $\psi(q) = 0$. Therefore, $\psi'(q) \ge 0$. Computing $\psi'(q)$ as above, the desired log-Sobolev inequality is derived.

□

Based on this theorem, in order prove hypercontractivity inequalities our main goal would be to estimate the 2-log-Sobolev constant α_2 which by definition is equal to

$$\alpha_2 = \inf_{f\ge 0}\frac{\mathcal{E}(f,f)}{\mathrm{Ent}_\pi(f^2)}.$$

Sometimes the 2-log-Sobolev inequality or the 2-log-Sobolev constant is referred just by log-Sobolev inequality or log-Sobolev constant.

In the following proposition we show that the 2-log-Sobolev constant provides a lower bound on the spectral gap.

Proposition 2.5. *$\lambda \ge 2\alpha_2$.*

Proof. Let g be a function with $\mathbb{E}g = 0$. Then for sufficiently small $|\epsilon| > 0$, the function $f_\epsilon = 1 + \epsilon g$ is positive and we have

$$\alpha_2\,\mathrm{Ent}_\pi(f_\epsilon^2) \le \mathcal{E}(f_\epsilon, f_\epsilon).$$

Therefore, $\psi(\epsilon) = \mathcal{E}(f_\epsilon, f_\epsilon) - \alpha_2\,\mathrm{Ent}(f_\epsilon^2) \ge 0$ is non-negative. On the other hand we have

$$\psi(\epsilon) = \epsilon^2\big(\mathcal{E}(g,g) - 2\alpha_2\mathbb{E}[g^2]\big) + O(\epsilon^3).$$

Therefore, $2\alpha_2\mathbb{E}g^2 \le \mathcal{E}(g,g)$. Comparing to (5) we obtain the desired bound on λ. □

Interestingly we can prove lower bounds on α_2 in terms of the spectral gap.

Theorem 2.6. *Suppose that for some $q \geq 2$ and $t_q > 0, M_q > 0$ we have*

$$\|T_{t_q}\|_{2\to q} \leq M_q.$$

Then we have

$$\alpha_2 \geq \frac{(1-2/q)\lambda}{2(\lambda t_q + \ln M_q + (q-2)/q)}$$

The proof of this theorem is based on interpolation theory for which we refer to [1, Theorem 3.9].

We now present the so-called *tensorization* property of log-Sobolev constants.

Theorem 2.7. *(Subadditivity of Entropy) Let $(\Omega_1 \times \Omega_2, \pi_1 \times \pi_2)$ be a product probability space and let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be an arbitrary non-negative function. Then we have*

$$\mathrm{Ent}_{\pi_1\times\pi_2} f \leq \mathbb{E}_{\pi_1}\big[\mathrm{Ent}_{\pi_2}(f)\big] + \mathbb{E}_{\pi_2}\big[\mathrm{Ent}_{\pi_1}(f)\big].$$

The proof of this theorem is based on the convexity of $s \mapsto s\log s$ and is left for the reader.

Theorem 2.8. *Let (Ω_k, π_k), for $k = 1, \dots, n$, be a probability spaces and let $\mathcal{L}_k$ be a Lindblad operator. Let $\alpha_q(\mathcal{L}_k)$ be the q-log-Sobolev constant of $\mathcal{L}_k$. Consider the product probability space $(\tilde{\Omega}, \tilde{\pi})$ with $\tilde{\Omega} = \Omega_1 \times \cdots \times \Omega_m$, and $\tilde{\pi} = \pi_1 \otimes \cdots \otimes \pi_m$. Then $\tilde{\mathcal{L}} = \hat{\mathcal{L}}_1 + \cdots \hat{\mathcal{L}}_m$, where $\hat{\mathcal{L}}_k$ is the lift of $\mathcal{L}_k$ acing on $L^2(\tilde{\pi})$, is a Lindblad operator which generates*

$$\tilde{T}_t = e^{-t\tilde{\mathcal{L}}} = e^{-t\mathcal{L}_1} \otimes \cdots \otimes e^{-t\mathcal{L}_m}.$$

Moreover we have

$$\alpha_q(\tilde{\mathcal{L}}) = \max_k \alpha_q(\mathcal{L}_k).$$

The proof of this theorem is a straightforward consequence of the sub-additivity of the entropy function.

3 Reverse hypercontractivity

In the previous section we derived inequalities of the form $\|T_t f\|_p \leq \|f\|_q$ for $p, q \geq 1$ using log-Sobolev inequalities. Interestingly log-Sobolev inequalities can also be used to find inequalities in the reverse direction $\|T_t f\|_p \geq \|f\|_q$ but for $p, q < 1$.

We first extend the definition of the p-norm for $p < 1$. For any positive function $f > 0$ and $p \neq 0$ define

$$\|f\|_p = (\mathbb{E} f^p)^{1/p},$$

as before, and for $p = 0$ let

$$\|f\|_0 = \lim_{p\to 0} \|f\|_p = e^{\mathbb{E}[\log f]}.$$

Observe that

$$\|f\|_{-p} = \|f^{-1}\|_p^{-1}. \tag{10}$$

$\|\cdot\|_p$ for $p<1$ does not satisfy triangle's inequality and is not a norm (but a semi-norm), yet by abuse of terminology we still call it a norm.

A variant of Hölder's duality still holds for $p<1$: for any $f>0$ we have

$$\|f\|_p = \inf_{g>0} \frac{\langle f, g\rangle}{\|g\|_{\hat{p}}},$$

where as before $\hat{p}$ is the Hölder conjugate of p given by (6) and $\hat{p}=0$ for $p=0$. Proposition 2.1 holds for $p<1$ as well and then the map $p \mapsto \|f\|_p$ is non-decreasing over $\mathbb{R}$ (not just for $p \geq 1$).

We define p-log-Sobolev inequalities for $p<1$ similarly as before according to (8) if $p \neq 0$. Log-Sobolev constants are defined similarly by (9). For $p=0$ the 0-log-Sobolev inequality would be obtained in the limit of $p \to 0$ and is given by

$$\alpha_0 \mathrm{Var}[\log f] \leq -\frac{1}{2}\mathcal{E}(f, 1/f).$$

Then again we have $\alpha_p = \alpha_{\hat{p}}$ for all p. Moreover, Theorem 2.3 extends to the interval $[0,2]$ by the same proof. Thus $p \mapsto \alpha_p$ is non-decreasing on $[0,2]$.

The tensorization property of log-Sobolev constants as stated in Theorem 2.8 is also generalized to α_p's for $p<1$ with the same proof.

Here we should mention that 0-log-Sobolev inequality is equivalent to the Poincare inequality

$$\lambda \mathrm{Var}[f] \leq \mathcal{E}(f,f), \qquad \forall f.$$

Proposition 3.1. *$\alpha_0 = \lambda/2$.*

Observe that by this proposition the inequality $\alpha_2 \leq \lambda/2$ that we proved in Proposition 2.5 follows from the monotonicity of log-Sobolev constants in Theorem 2.3.

Proof. The proof of $\lambda \geq 2\alpha_0$ is similar to that of Proposition 2.5, i.e., consider the 0-log-Sobolev inequality for $f = 1+\epsilon g$ for sufficiently small $|\epsilon|>0$. The other direction $2\alpha_0 \geq \lambda$ is derived from the inequality

$$\mathcal{E}(\log f, \log f) \leq -\mathcal{E}(f, 1/f),$$

for whose proof we refer to [12, Corollary 2.5]. □

By a convexity type argument we can again show that for any $p<1$ we have

$$\|T_t f\|_p \geq \|f\|_p, \qquad \forall f>0.$$

That is, T_t is a *reverse contraction* under p-norms for all $p<1$. However, using p-log-Sobolev inequalities for $p<1$ we can prove stronger inequalities that are called *reverse hypercontractivity inequalities.*

Theorem 3.2. *For every $f > 0$ we have*

$$\|T_t f\|_p \geq \|f\|_q, \qquad \forall p < q < 1, \quad e^{4\alpha_1 t} \geq \frac{1-p}{1-q}.$$

The proof of this theorem is identical to that of Theorem 2.4. The only difference is that here since we are interested in $p < q < 1$ the relevant log-Sobolev constants α_p are those with $p \in [0, 1]$. The smallest such log-Sobolev constant is α_1 and this is why α_1 appears in the statement of theorem (instead of α_2 as compared to Theorem 2.4).

This theorem shows that 1-log-Sobolev inequalities are also special (besides 2-log-Sobolev inequalities) and sometimes they are called *modified log-Sobolev inequalities.* Modified log-Sobolev inequalities are important not only for proving reverse hypercontractivity inequalities, but also for proving bounds on the entropy production of Markov semigroups.

Proposition 3.3. *For any reversible Markov semigroup $\{T_t : t \geq 0\}$ and $f \geq 0$ we have*

$$\mathrm{Ent}_\pi(T_t f) \leq e^{-4\alpha_1 t}\mathrm{Ent}_\pi(f).$$

Proof. Using $\mathbb{E}[T_t f] = \mathbb{E}f$ and the 1-log-Sobolev inequality we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}_\pi(T_t f) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[T_t f \log T_t f] = -\mathbb{E}[\mathcal{L}T_t f \log T_t f] = -\mathcal{E}(T_t f, \log T_t f) \leq -4\alpha_1 \mathrm{Ent}_\pi(T_t f).$$

Integrating this inequality we obtain the desired result.

□

Let us finish this section by mentioning that hypercontractivity inequalities were first studied in the context of quantum field theory in [19, 20]. Also, the notion of log-Sobolev inequalities was first introduced in the seminal paper of Gross [18]. For a more detailed history of the subject we refer the reader to [24]. For the history of reverse hypercontractivity inequalities see [12] and reference therein.

4 From discrete to continuous Markov processes

In this section we introduce an interesting example of Markov semigroups. Let K be a stochastic matrix. For instance, K may be the transition matrix of a random walk on a graph. Then the transition matrix at the m-th step of the walk is K^m, and $\{K^m : m \in \mathbb{N}\}$ is a (discrete) semigroup. In the following we associate a continuous Markov semigroup to K.

To obtain a continuous process, instead of making a move at each time step, we may wait for a random amount of time and then make a move according to the transition matrix K. To get a Markov process, this waiting time must be memoryless, i.e., the waiting time must come from an exponential random variable . In other words, our moves are going to be made at points determined by a Poisson process. To describe this process more precisely, let N_t be an independent Poisson process with rate t. Then we define

$$T_t = \mathbb{E}[K^{N_t}],$$

where the expectation is over N_t. Indeed we have

$$T_t = e^{-t}\sum_{j=0}^{\infty}\frac{t^j}{j!}K^j.$$

From the above equation it is clear that $T_t = e^{-t\mathcal{L}}$ where

$$\mathcal{L} = I - K.$$

Assuming that K is reversible with stationary distribution π, this Lindblad operator $\mathcal{L}$ is self-adjoint as an operator acting on $L^2(\pi)$. In this case, we have

$$\begin{aligned}
\mathcal{E}(f,g) &= \langle f,(I-K)g\rangle_\pi \\
&= \sum_x \pi(x)f(x)\Big(g(x) - \sum_y K(x,y)g(y)\Big) \\
&= \sum_x \pi(x)f(x)\sum_y \Big(K(x,y)(g(x)-g(y))\Big) \\
&= \sum_{x,y}\pi(x)K(x,y)f(x)(g(x)-g(y)).
\end{aligned}$$

where in the third line we use $K1 = 1$. Using the detailed balance condition (3) we can also write

$$\begin{aligned}
\mathcal{E}(f,g) &= \sum_{x,y}\pi(y)K(y,x)f(x)(g(x)-g(y)) \\
&= \sum_{x,y}\pi(x)K(x,y)f(y)(g(y)-g(x)).
\end{aligned}$$

Then taking the average of the above two equations we find that

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{x,y}K(x,y)\pi(x)(f(x)-f(y))(g(x)-g(y)). \tag{11}$$

From this formula it is clear that $\mathcal{E}(f,f)\geq 0$ which was proved before.

Example 1: Let K be the trivial transition matrix with $K(x,y) = \pi(y)$ where π is an arbitrary distribution. Then we have $Kf = \mathbb{E}f$, and $\mathcal{L} = I - \mathbb{E}$ is the example that we considered in Section 1 as well. $\mathcal{L} = I - \mathbb{E}$ is sometimes called the *simple generator*. In this case using $\mathbb{E}^2 = \mathbb{E}$ we have

$$T_t = e^{-t}I + (1-e^{-t})\mathbb{E}.$$

We also have

$$\mathcal{E}(f,f) = \mathrm{Var}f.$$

The spectral gap is easily verified to be equal to $\lambda = 1$. It is shown in [1, Theorem A.1] that

$$\alpha_2 = \frac{1 - 2\pi_{\min}}{\log(\pi_{\min}^{-1} - 1)},$$

where

$$\pi_{\min} = \min_x \pi(x).$$

In particular, if π is the uniform distribution we have

$$\alpha_2 = \frac{1 - \frac{2}{|\Omega|}}{\log(|\Omega| - 1)},$$

and if $|\Omega| = 2$ we have $\alpha_2 = 1/2$. We will study this latter case in more detailes later on.

Estimating α_1 for this Markov semigroup is easy. Using the concavity of the logarithm function we have

$$\begin{aligned}
\mathrm{Ent}_\pi(f) &= \mathbb{E}[f \log f] - \mathbb{E}[f] \log \mathbb{E}[f] \\
&\leq \mathbb{E}[f \log f] - \mathbb{E}[f]\mathbb{E}[\log f] \\
&= \mathbb{E}[f(\log f - \mathbb{E} \log f)] \\
&= \mathcal{E}(f, \log f).
\end{aligned}$$

Therefore, considering the normalization factor of the 1-log-Sobolev inequality we have $\alpha_1 \geq 1/4$.

The following corollary first appeared in [1, Corollary A.4] gives a general bound on the 2-log-Sobolev constant.

Corollary 4.1. *For any Lindblad operator $\mathcal{L}$ with spectral gap λ we have*

$$\alpha_2 \geq \frac{(1 - 2\pi_{\min})\lambda}{\log(\pi_{\min}^{-1} - 1)}$$

Proof. Having the log-Sobolev constant of the previous example we have

$$\frac{1 - 2\pi_{\min}}{\log(\pi_{\min}^{-1} - 1)} \mathrm{Ent}(f^2) \leq \langle f, (I - \mathbb{E})f \rangle_\pi = \mathrm{Var} f.$$

Combining this with the Poincare inequality $\lambda \mathrm{Var} f \leq \mathcal{E}(f, f)$ we obtain the desired result. □

Example 2: Let K denote the random walk on the complete graph on n vertices. Then π is the uniform distribution $\pi = 1/n$, and we have [1]

$$\alpha_2 = \frac{n - 2}{(n - 1)\log(n - 1)},$$

if $n > 3$, and for $n = 2$ we have $\alpha_2 = 1$. Observe that the inequality of Corollary 4.1 is tight for the example of complete graph. The following bounds on the 1-log-Sobolev constant of K are derived in [3, Lemma 2.6]

$$\frac{n}{4(n-1)} \leq \alpha_1 \leq \Big(\frac{1}{4} + \frac{1}{\log(n+1)}\Big)\frac{n}{n-1}$$

Example 3: Let K denote the random walk on the hypercube $\Omega = \{+1, -1\}^n$ where $x, y \in \Omega$ are adjacent if they are different only in a single coordinate. The stationary distribution of this random walk is the uniform distribution $\pi = 1/2^n$. Define

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A is associated with the random walk on the complete graph on 2 vertices. Then we have

$$K = \frac{1}{n}(\hat{A}_1 + \ldots \hat{A}_n),$$

where $\hat{A}_j$ is the lift of A acting on the j-th coordinate. We saw in the previous example that $\alpha_2(I - A) = 1$. Then using Theorem 2.8 we have

$$\alpha_2(I - K) = \frac{1}{n}\alpha_2(I - A) = \frac{1}{n}.$$

Note that in this case $\lambda = 2/n$.

Example 4: Let $\Omega = S_n$ be the symmetric group. Consider the random walk on S_n with transition matrix $K(\sigma, \sigma') = 2/(n(n-1))$ if $\sigma^{-1}\sigma'$ is a transposition, and $K(\sigma, \sigma') = 0$ otherwise. The stationary distribution is $\pi = 1/n!$. We have $\lambda = 2/(n-1)$. It is proved in [1] that

$$\frac{1}{3n \ln n} \leq \alpha_2 \leq \frac{1}{n-1}.$$

Also it is shown in [3, Corollary 3.1] that

$$\frac{1}{4(n-1)} \leq \alpha_1 \leq \frac{1}{n-1}.$$

5 The Ornstein-Uhlenbeck semigroup

So far we have considered only discrete probability spaces to avoid some technicalities that arise in the continue case. Nevertheless, the theory of log-Sobolev inequalities and hypercontractivity inequalities is developed similarly in the continue case as well. In this section we study the important example of the *Ornstein-Uhlenbeck semigroup.*

Recall that the density of the normal (Gaussian) distribution $\mathcal{N}(a, \sigma^2)$ on $\mathbb{R}$ with mean a and variance σ^2 is given by

$$\mathrm{d}\nu = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} \mathrm{d}x.$$

That is, for every measurable (and sufficiently nice) function $f : \mathbb{R} \to \mathbb{R}$ we have

$$\mathbb{E}_\nu[f] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} f e^{-\frac{(x-a)^2}{2\sigma^2}} \mathrm{d}x.$$

We are in particular interested in the normal distribution $\mathcal{N}(0, 1)$ with zero-mean and variance 1 which we call the *standard normal distribution.* The importance of the normal distribution relies on the famous central limit theorem.

Theorem 5.1. *(Central limit theorem) Let $X_1, \ldots, X_n$ be i.i.d. real random variables with mean $\mathbb{E}[X] = a$ and variance $\mathrm{Var}[X] = \sigma^2$. Then*

$$S_n := \frac{X_1 + \cdots + X_n - na}{\sqrt{n}},$$

converges to the normal distribution $\mathcal{N}(0, \sigma^2)$ in distribution as $n \to \infty$. That is,

$$\lim_{n\to\infty} \Pr[S_n \leq s] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{s} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x$$

Here is a simple consequence of the central limit theorem that can also be proven directly as an exercise.

Proposition 5.2. *Let X_i, for $i = 1, \ldots, k$ be independent normal random variables with distributions $\mathcal{N}(a_i, \sigma_i^2)$. Then for arbitrary constants $c_1, \ldots, c_k$, the random variable $\sum_i c_i X_i$ is also normal with distribution $\mathcal{N}(a, \sigma)$ where $a = \sum_i c_i a_i$ and $\sigma^2 = \sum_i c_i^2 \sigma_i^2$.*

For any sufficiently nice function $f : \mathbb{R} \to \mathbb{R}$ define the operator $\mathcal{L}$ by

$$\mathcal{L}f(x) := xf'(x) - f''(x).$$

We are going to show that $\mathcal{L}$ is a valid Lindblad generator of a Markov semigroup and compute the corresponding semigroup.

In the following, all expectations are with respect to the standard normal distribution $\pi = \mathcal{N}(0, 1)$. In particular, for two real functions f, g we denote

$$\langle f, g\rangle_\pi = \mathbb{E}[fg] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)g(x)e^{-\frac{x^2}{2}} \mathrm{d}x.$$

Lemma 5.3. *(a) $\mathcal{L}1 = 0$.*

(b) $\mathbb{E}[\mathcal{L}f] = \mathbb{E}[xf' - f''] = 0$ for all f.

(c) $\langle f, \mathcal{L}g\rangle_\pi = \mathbb{E}[f'g'] = \langle \mathcal{L}f, g\rangle_\pi$.

Note that (c) shows that $\mathcal{L}$ is reversible with respect to the standard normal distribution $\pi = \mathcal{N}(0, 1)$.

Proof. Item (a) is obvious from the definition. Item (b) follows from (a) and (c):

$$\mathbb{E}[\mathcal{L}f] = \langle 1, \mathcal{L}f\rangle_\pi = \langle \mathcal{L}1, f\rangle_\pi = \langle 0, f\rangle_\pi = 0.$$

To prove (c) we use integration by parts:

$$\begin{aligned}
\langle f, \mathcal{L}g\rangle_\pi &= \langle f, xg' - g''\rangle_\pi \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)\big(xg'(x) - g''(x)\big)e^{-\frac{x^2}{2}} \mathrm{d}x \\
&= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)\Big(g'(x)e^{-\frac{x^2}{2}}\Big)' \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(x)g'(x)e^{-\frac{x^2}{2}} \mathrm{d}x \\
&= \mathbb{E}[f'g'].
\end{aligned}$$

$\square$

We now compute the semigroup generated by $\mathcal{L}$. Define

$$T_t f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f\big(e^{-t}x + \sqrt{1-e^{-2t}}\, y\big) e^{-\frac{y^2}{2}} \mathrm{d}y. \tag{12}$$

Observe that $T_t f(x)$ can equivalently be written as

$$T_t f(x) = \mathbb{E}_Y\Big[f\big(e^{-t}x + \sqrt{1-e^{-2t}}\, Y\big)\Big],$$

with Y being a standard normal random variable.

The following proposition shows that $\{T_t : t \geq 0\}$ forms a semigroup whose generator is $\mathcal{L}$. This semigroup is called the Ornstein-Uhlenbeck semigroup.

Proposition 5.4. *(a) $T_t f \geq 0$ whenever $f \geq 0$, and $T_t 1 = 1$.*

(b) $\lim_{t\to 0^+} T_t f(x) = f(x)$ and $\lim_{t\to\infty} T_t f(x) = \mathbb{E}[f]$ for every bounded continuous function f.

(c) $T_t T_s = T_{s+t}$.

(d) $T_t \mathcal{L} = \mathcal{L} T_t$

(e) We have

$$\frac{\partial}{\partial t} T_t f(x) = -\mathcal{L} T_t f(x).$$

and then $\mathbb{E}[T_t f] = \mathbb{E}[f]$.

Proof. (a) is obvious from the definition and (b) is left as an exercise. To prove (c) we compute

$$\begin{aligned} T_t T_s f(x) &= \mathbb{E}_Y\Big[T_s f\big(e^{-t}x + \sqrt{1-e^{-2t}}\, Y\big)\Big] \\ &= \mathbb{E}_Y \mathbb{E}_Z\Big[f\Big(e^{-s}\big(e^{-t}x + \sqrt{1-e^{-2t}}\, Y\big) + \sqrt{1-e^{-2s}} Z\Big)\Big] \\ &= \mathbb{E}_Y \mathbb{E}_Z\Big[f\Big(e^{-(s+t)}x + e^{-s}\sqrt{1-e^{-2t}} Y + \sqrt{1-e^{-2s}} Z\Big)\Big], \end{aligned}$$

where Y, Z are independent standard normal distributions. Now using Proposition 5.2, the random variable $e^{-s}\sqrt{1-e^{-2t}}Y + \sqrt{1-e^{-2s}}Z$ is also normal with distribution $\mathcal{N}(0, \sigma^2)$ where

$$\sigma^2 = e^{-2s}(1-e^{-2t}) + (1-e^{-2s}) = 1 - e^{-2(s+t)}.$$

Therefore, for a standard normal U we have

$$T_t T_s f(x) = \mathbb{E}_U\Big[f\big(e^{-(s+t)}x + \sqrt{1-e^{-2(s+t)}}\, U\big)\Big] = T_{s+t} f(x).$$

To prove (d) let us define $g(y) := f(e^{-t}x + \sqrt{1-e^{-2t}}\,y)$. Then using $\mathbb{E}[\mathcal{L}g] = 0$ we have

$$\begin{aligned}
T_t\mathcal{L}f(x) &= \mathbb{E}_Y\Big[\mathcal{L}f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\Big] \\
&= \mathbb{E}_Y\Big[\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big) - f''\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\Big] \\
&= e^{-t}x\mathbb{E}_Y\big[f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] - e^{-2t}\mathbb{E}\big[f''\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] \\
&\qquad + \mathbb{E}_Y[Yg'(Y) - g''(Y)] \\
&= e^{-t}x\mathbb{E}_Y\big[f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] - e^{-2t}\mathbb{E}_Y\big[f''\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] \\
&= x\frac{\partial}{\partial x}\mathbb{E}_Y\big[f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] - \frac{\partial^2}{\partial x^2}\mathbb{E}_Y\big[f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] \\
&= x\frac{\partial}{\partial x}T_tf(x) - \frac{\partial^2}{\partial x^2}T_tf(x) \\
&= \mathcal{L}T_tf(x).
\end{aligned}$$

For (e) we again use $\mathbb{E}_Y[Yg'(Y)] = \mathbb{E}_Y[g''(Y)]$ for the above function g as follows:

$$\begin{aligned}
\frac{\partial}{\partial t}T_tf(x) &= \frac{\partial}{\partial t}\mathbb{E}_Y\big[f\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] \\
&= \mathbb{E}_Y\Big[\Big(-e^{-t}x + \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\,Y\Big)f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\Big] \\
&= -e^{-t}x\mathbb{E}_Y\big[f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] + \frac{e^{-2t}}{1-e^{-2t}}\mathbb{E}_Y[Yg'(Y)] \\
&= -e^{-t}x\mathbb{E}_Y\big[f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] + \frac{e^{-2t}}{1-e^{-2t}}\mathbb{E}_Y[g''(Y)] \\
&= -e^{-t}x\mathbb{E}_Y\big[f'\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)\big] + e^{-2t}\mathbb{E}_Y[f''\big(e^{-t}x + \sqrt{1-e^{-2t}}\,Y\big)] \\
&= -\mathcal{L}T_tf(x),
\end{aligned}$$

where the last step follows from the computations done above. □

Now that we have a Markov semigroup, it is natural to estimate its log-Sobolev constants.

Theorem 5.5. *Let $\mathcal{L}f = xf' - f''$ be the generator of the Ornstein-Uhlenbeck semigroup. Then we have*

$$\alpha_2(\mathcal{L}) \geq \frac{1}{2}.$$

Equivalently, for every function f we have

$$\mathrm{Ent}_\pi(f^2) \leq 2\langle f, \mathcal{L}f\rangle_\pi = 2\mathbb{E}[f'^2] = \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f'(x)^2 e^{-\frac{x^2}{2}}\,\mathrm{d}x. \tag{13}$$

Note that the entropy function here is defined as before:

$$\mathrm{Ent}_\pi(g) = \mathbb{E}[g\log g] - \mathbb{E}[g]\log\mathbb{E}[g],$$

where the expectations are with respect to the standard normal distribution.

Proof. We will present two proofs for this theorem.

First proof: For a smooth bounded function g we have $T_0 g = g$ and $\lim_{t\to\infty} T_t g = \mathbb{E}[g]$. Therefore, letting $g = f^2$ we have

$$\begin{aligned}
\mathrm{Ent}_\pi(f^2) &= \mathbb{E}[g \log g] - \mathbb{E}[g] \log \mathbb{E}[g] \\
&= -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\big[T_t g \cdot \log T_t g\big] \mathrm{d}t \\
&= -\int_0^\infty -\mathbb{E}\big[\mathcal{L} T_t g \cdot \log T_t g - \mathcal{L} T_t(g)\big] \mathrm{d}t \\
&= \int_0^\infty \mathbb{E}\big[\mathcal{L} T_t g \cdot \log T_t g\big] \mathrm{d}t \\
&= \int_0^\infty \langle \mathcal{L} T_t g, \log T_t g\rangle \, \mathrm{d}t \\
&= \int_0^\infty \mathbb{E}\big[(T_t g)' \cdot (\log T_t g)'\big] \mathrm{d}t \\
&= \int_0^\infty \mathbb{E}\Big[\frac{(T_t g)'^2}{T_t g}\Big] \mathrm{d}t
\end{aligned}$$

We have

$$T_t g(x) = \mathbb{E}_Y\big[g(e^{-t} x + \sqrt{1 - e^{-2t}}\, Y)\big].$$

Therefore, using the Cauchy-Schwarz inequality we have

$$\begin{aligned}
(T_t g)'(x) &= e^{-t} \mathbb{E}_Y\big[g'(e^{-t} x + \sqrt{1 - e^{-2t}}\, Y)\big] \\
&= 2e^{-t} \mathbb{E}_Y\big[f(e^{-t} x + \sqrt{1 - e^{-2t}}\, Y) f'(e^{-t} x + \sqrt{1 - e^{-2t}}\, Y)\big] \\
&\le 2e^{-t} \Big(\mathbb{E}_Y\big[f^2(e^{-t} x + \sqrt{1 - e^{-2t}}\, Y)\big] \mathbb{E}_Y\big[f'^2(e^{-t} x + \sqrt{1 - e^{-2t}}\, Y)\big]\Big)^{1/2} \\
&= 2e^{-t} \Big(T_t g(x) T_t(f'^2)(x)\Big)^{1/2}.
\end{aligned}$$

We conclude that

$$\frac{(T_t g)'(x)^2}{T_t g(x)} \le 4e^{-2t} T_t(f'^2)(x).$$

As a result,

$$\begin{aligned}
\mathrm{Ent}_\pi(f^2) &\le \int_0^\infty 4e^{-2t} \mathbb{E}\big[T_t(f'^2)\big] \mathrm{d}t \\
&= \int_0^\infty 4e^{-2t} \mathbb{E}\big[f'^2\big] \mathrm{d}t \\
&= 2\mathbb{E}\big[f'^2\big].
\end{aligned}$$

Second proof: Observe that by Example 1, the 2-log-Sobolev constant in this theorem equals the 2-log-Sobolev constant of the simple Lindblad operator $I - \mathbb{E}$ associated with the uniform distribution on $\Omega = \{+1, -1\}$. This together with the central limit theorem gives another proof of the above log-Sobolev inequality [18].

Using the tensorization property of log-Sobolev inequalities, i.e., Theorem 2.8, together with the fact that $\alpha_2(I-\mathbb{E}) = 1/2$ we find that for every function $g : \{+1,-1\}^n \to \mathbb{R}$

$$\mathrm{Ent}(g^2) \leq \frac{1}{2}\mathbb{E}\Big[\sum_{i=1}^n \big(g(x_1,\dots,x_n) - g(x_1,\dots,-x_i,\dots,x_n)\big)^2\Big]. \tag{14}$$

Let us take an arbitrary $f : \mathbb{R} \to \mathbb{R}$, say with bounded first and second derivatives, and in the above inequality let

$$g_n(x_1,\dots,x_n) := f\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big).$$

By the central limit theorem the distribution of $\frac{x_1+\cdots+x_n}{\sqrt{n}}$, when $(x_1,\dots,x_n)$ is chosen uniformly in $\{+1,-1\}^n$, converges to the standard normal distribution. Therefore, $\mathrm{Ent}(g_n^2)$ converges to $\mathrm{Ent}_\pi(f^2)$ as $n \to \infty$. Analyzing the limit of right hand side needs more work. Observer that

$$\begin{aligned} g_n(x_1,\dots,x_n) - g_n(x_1,\dots,-x_i,\dots,x_n) &= f\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big) - f\big(\frac{x_1+\cdots+x_n-2x_i}{\sqrt{n}}\big) \\ &= \frac{2x_i}{\sqrt{n}} f'\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big) + O\Big(\Big|\frac{2x_i}{\sqrt{n}}\Big|^2\Big) \\ &= \frac{2x_i}{\sqrt{n}} f'\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big) + O\Big(\frac{1}{n}\Big). \end{aligned}$$

Therefore,

$$\big(g(x_1,\dots,x_n) - g(x_1,\dots,-x_i,\dots,x_n)\big)^2 = \frac{4}{n} f'^2\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big) + O\Big(\frac{1}{n^{3/2}}\Big),$$

and

$$\sum_{i=1}^n \big(g(x_1,\dots,x_n) - g(x_1,\dots,-x_i,\dots,x_n)\big)^2 = 4f'^2\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big) + O\Big(\frac{1}{\sqrt{n}}\Big).$$

Now as before, using the central limit theorem we find that

$$\mathbb{E}\Big[f'^2\big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\big)\Big]$$

tends to $\mathbb{E}[f'^2]$ as $n \to \infty$ and the $O(1/n)$ term vanishes. Using these in (14) we obtain the desired inequality. □

The above theorem can be applied to functions f that are not necessarily smooth. It suffices to replace f' on the hand side of (13) with

$$f'(x) := \limsup_{y \to x} \frac{f(y) - f(x)}{y - x}.$$

In particular, (13) holds for Lipschitz functions as well.

By the Stroock-Varopoulos inequality we have $\alpha_1(\mathcal{L}) \geq \alpha_2(\mathcal{L}) = 1/2$. Therefore, for every g we have

$$\mathrm{Ent}_\pi(e^g) \leq \frac{1}{2}\langle g, \mathcal{L}e^g\rangle = \frac{1}{2}\mathbb{E}[g'^2 e^g]. \tag{15}$$

Another way to see this (without using the Stroock-Varopoulos inequality) is to use (13) for the function $f = e^{g/2}$.

The Ornstein-Uhlenbeck semigroup can also be defined in higher dimensions. Let us first recall some notations. Vectors in $\mathbb{R}^k$ are denoted by boldface letters $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$. We also use the notation

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{k} x_i y_i,$$

and $|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x} = \sum_i x_i^2$. Then the gradient of a function $f : \mathbb{R}^k \to \mathbb{R}$ equals

$$\nabla f = (\partial_1 f, \cdots, \partial_k f),$$

where for simplicity of notation we use

$$\partial_i = \frac{\partial}{\partial x_i}.$$

Then we have

$$|\nabla f|^2 = \sum_{i=1}^{k} (\partial_i f)^2.$$

The divergence of a tuple of functions $(f_1, \ldots, f_k)$ is defined by

$$\mathrm{div}(f_1, \ldots, f_k) = \sum_{i}^{k} \partial_i f_i.$$

Then $\mathrm{div}\nabla f$ equals the Laplacian:

$$\Delta f = \mathrm{div}(\nabla f) = \sum_{i=1}^{k} \partial_i^2 f.$$

Corollary 5.6. *For every sufficiently nice function $f : \mathbb{R}^k \to \mathbb{R}$ we have*

$$\mathrm{Ent}_\pi(f^2) \leq \frac{2}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} |\nabla f(\mathbf{x})|^2 e^{-\frac{|\mathbf{x}|^2}{2}} \mathrm{d}\mathbf{x}.$$

Here the entropy function is defined with respect to the k-dimensional standard normal distribution, i.e., when the expectations are given by

$$\mathbb{E}[g] = \frac{2}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} g(\mathbf{x}) e^{-\frac{|\mathbf{x}|^2}{2}} \mathrm{d}\mathbf{x}.$$

Proof. This corollary follows from the tensorization property of log-Sobolev inequalities, i.e., Theorem 2.8. The point is that product of Gaussian distributions becomes a *multidimensional* Gaussian distribution. In particular, the density of the product of n standard normal distributions is

$$\prod_{i=1}^{k}\Big(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_i^2}{2}}\Big)\mathrm{d}x_i = \frac{1}{(2\pi)^{\frac{k}{2}}}e^{-\frac{|\mathbf{x}|^2}{2}}\mathrm{d}\mathbf{x}. \tag{16}$$

Moreover, the Lindblad operator $\tilde{\mathcal{L}} = \hat{\mathcal{L}}_1 + \cdots + \hat{\mathcal{L}}_n$ is given by

$$\tilde{\mathcal{L}}f(\mathbf{x}) = \sum_{i=1}^{k} x_i\partial_i f(\mathbf{x}) - \partial_i^2 f(\mathbf{x}) = \mathbf{x}\cdot\nabla f(\mathbf{x}) - \Delta f(\mathbf{x}).$$

It is not hard to verify that the Dirichlet form associated with $\tilde{\mathcal{L}}$ is equal to

$$\langle f, \tilde{\mathcal{L}}g\rangle_\pi = \mathbb{E}[\nabla f\cdot\nabla g],$$

where the expectation is with respect to the density (16). Using these in the associated log-Sobolev inequality that is derived from Theorem 5.5 and Theorem 2.8 we obtain the desired result. □

We finish this section with the remark that the above computations for the standard normal distribution can be generalized to other distributions on $\mathbb{R}^k$. Let π be a distribution on $\mathbb{R}^k$ with density

$$\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x.$$

Then define the Lindblad operator $\mathcal{L} = \nabla V\cdot\nabla - \Delta$ by

$$\mathcal{L}f(\mathbf{x}) = \sum_{i=1}^{k}\partial_i V(\mathbf{x})\partial_i f(\mathbf{x}) - \partial_i^2 f(\mathbf{x}). \tag{17}$$

It can be shown that this Lindblad operator is reversible with respect to π, and whose associated Dirichlet form is

$$\langle f, \mathcal{L}g\rangle_\pi = \int_{\mathbb{R}^k}\nabla f\cdot\nabla g\mathrm{d}\pi = \int_{\mathbb{R}^k} e^{-V(\mathbf{x})}\nabla f\cdot\nabla g\mathrm{d}\mathbf{x}$$

Observe that for the standard normal distribution we have $V(\mathbf{x}) = |\mathbf{x}|^2/2 + k\ln(2\pi)/2$ in which case (17) reduces to what we had before.

The log-Sobolev inequality associated to such distributions on $\mathbb{R}^k$ can be stated in terms of the so-called *Bakry-Emery criterion* [21].

Theorem 5.7. *Let π be a Borel probability measure on $\mathbb{R}^k$ with $\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x$ such that $\mathrm{Hess}(V)\geq cI$ for some constant c. Then for all functions f we have*

$$\mathrm{Ent}_\pi(f^2)\leq\frac{2}{c}\int|\nabla f|^2\mathrm{d}\pi.$$

Equivalently, we have $\alpha_2(\nabla V\cdot\nabla - \Delta)\geq c/2$.

Observe that for the standard Gaussian measure $V(\mathbf{x}) = -\frac{1}{2}|\mathbf{x}|^2$ and $\mathrm{Hess}(V) = I$, in which case the above theorem coincides with the log-Sobolev inequality for the standard normal distribution.

The above theorem can be proven based on similar ideas as in the first proof of Theorem 5.5 (See Appendix A). Later, we will give yet another proof of this theorem.

6 Bounding the mixing time

Let (Ω, π) be a probability space and let $\mathcal{L}$ be a primitive reversible Lindblad operator with positive spectral gap $\lambda > 0$. As we argued in the remark following Proposition 1.2, for any other probability measure μ on Ω we have $\mu T_t \to \pi$ as t tends to ∞. Our goal in this section is to derive bounds on the rate of this convergence. To this end let us define

$$\tau_{\text{mix}} := \min\{t : \ \|\mu T_t - \pi\|_{\text{TV}} \leq \frac{1}{2e}, \ \forall \mu\}, \tag{18}$$

where $\|\cdot\|_{\text{TV}}$ denotes the *total variation distance* defined by

$$\|\rho\|_{\text{TV}} = \frac{1}{2}\sum_x |\rho(x)|.$$

The choice of constant $1/(2e)$ in the definition of τ_{mix} is arbitrary. To see this, suppose that $\|\mu T_t - \pi\|_{\text{TV}} \leq 1/(2e)$ for all measures μ. Let ρ be an arbitrary function with $\sum_x \rho(x) = 0$. Let $\rho_+(x) = \max\{\rho(x), 0\}$, and $\rho_- = \rho_+ - \rho$. Then $\tilde{\rho}_+ = \rho_+/\theta$ and $\tilde{\rho}_- = \rho_-/\theta$ for $\theta = \|\rho\|_{\text{TV}}$ are probability distributions. We compute

$$\begin{aligned}
\|\rho T_t\|_{\text{TV}} &= \|\rho_+ T_t - \rho_- T_t\|_{\text{TV}} \\
&= \theta \|\tilde{\rho}_+ T_t - \tilde{\rho}_- T_t\|_{\text{TV}} \\
&\leq \theta\Big(\|\tilde{\rho}_+ T_t - \pi\|_{\text{TV}} + \|\tilde{\rho}_- T_t - \pi\|_{\text{TV}}\Big) \\
&\leq \|\rho\|_{\text{TV}}/e.
\end{aligned}$$

Now let μ be an arbitrary measure. Since $\sum_x(\mu T_t - \pi)(x) = 0$ we have

$$\|\mu T_{2t} - \pi\|_{\text{TV}} = \|(\mu T_t - \pi)T_t\|_{\text{TV}} \leq \|\mu T_t - \pi\|_{\text{TV}}/e \leq 1/2e^2 \leq 1/e^2.$$

Similarly, by a simple induction we have

$$\|\mu T_{mt} - \pi\|_{\text{TV}} \leq 1/e^m.$$

This means that a bound on the mixing time τ_{mix} for error $1/e$ gives the bound $\tau_{\text{mix}} \log(\epsilon^{-1})$ on the mixing time for error ϵ.

Let us turn back to the problem of estimating the mixing time τ_{mix} defined in (18). Our first approach for finding an upper bound on τ_{mix} is via Proposition 1.2. For this we first need to upper bound the total variation distance in terms of the 2-norm. For any probability measure μ we have

$$\begin{aligned}
\|\mu-\pi\|_{\mathrm{TV}} &= \max_{f:f(x)\in\{\pm 1\}} \frac{1}{2}\sum_x f(x)\cdot(\mu(x)-\pi(x)) \\
&= \max_{f:f(x)\in\{\pm 1\}} \frac{1}{2}\sum_x f(x)\pi(x)^{1/2}\cdot(\mu(x)-\pi(x))\pi(x)^{-1/2} \\
&\leq \max_{f:f(x)\in\{\pm 1\}} \frac{1}{2}\sqrt{\sum_x f(x)^2\pi(x)}\cdot\sqrt{\sum_x(\mu(x)-\pi(x))^2\pi(x)^{-1}} \\
&= \frac{1}{2}\sqrt{\sum_x(\mu(x)-\pi(x))^2\pi(x)^{-1}} \\
&= \frac{1}{2}\Big\|\frac{\mu}{\pi}-1\Big\|_2,
\end{aligned}$$

where by $\frac{\mu}{\pi}$ we mean the function $\frac{\mu}{\pi}(x)=\mu(x)/\pi(x)$. Using this inequality we have

$$\|\mu T_t-\pi\|_{\mathrm{TV}}\leq\frac{1}{2}\Big\|\frac{\mu T_t}{\pi}-1\Big\|_2.$$

Now using (4) based on the detailed balance condition we have

$$\frac{\mu T_t}{\pi}=T_t f,$$

where $f=\mu/\pi$. We continue

$$\|\mu T_t-\pi\|_{\mathrm{TV}}\leq\frac{1}{2}\|T_t f-1\|_2=\frac{1}{2}\|T_t f-\mathbb{E}f\|_2.$$

As a result, by Proposition 1.2 we have

$$\|\mu T_t-\pi\|_{\mathrm{TV}}\leq\frac{1}{2}e^{-\lambda t}\|f-1\|_2.$$

Therefore,

$$\begin{aligned}
\max_{\mu}\|\mu T_t-\pi\|_{\mathrm{TV}} &\leq \frac{1}{2}e^{-\lambda t}\max_{f:f\geq 0,\mathbb{E}f=1}\|f-1\|_2 \\
&= \frac{1}{2}e^{-\lambda t}\sqrt{\pi_{\min}\Big(\frac{1}{\pi_{\min}}-1\Big)^2} \\
&\leq \frac{1}{2}\sqrt{\frac{1}{\pi_{\min}}}e^{-\lambda t},
\end{aligned}$$

where

$$\pi_{\min}=\min_x\pi(x).$$

Putting all these together we arrive at the following theorem.

Theorem 6.1. *For every reversible Markov semigroup we have*

$$\tau_{\text{mix}} \leq \frac{1}{\lambda}\Big(1 + \frac{1}{2}\log\frac{1}{\pi_{\min}}\Big).$$

We now argue that the bound given by the above theorem is loose when α_1 and λ are of the same order.

Theorem 6.2. *For every reversible Markov semigroup we have*

$$\tau_{\text{mix}} \leq \frac{1}{4\alpha_1}(\log 2 + 2 + \log\log\frac{1}{\pi_{\min}}) \leq \frac{1}{4\alpha_2}(\log 2 + 2 + \log\log\frac{1}{\pi_{\min}}).$$

In the proof of this theorem we use *Pinsker's inequality*

$$\|\mu - \pi\|_{\text{TV}}^2 \leq \frac{1}{2}D(\mu\|\pi), \tag{19}$$

where $D(\mu\|\pi)$ is the KL-divergence defined in (7). Indeed this inequality will be replaced with the estimation of the total variation distance with the 2-norm (sometimes called the χ^2-divergence) in the proof of Theorem 6.1.

Proof. Let μ be an arbitrary probability measure. Then for $f_t = \frac{\mu T_t}{\pi}$ we have $f_t = T_t f_0$. We then have

$$\|\mu T_t - \pi\|_{\text{TV}}^2 \leq \frac{1}{2}D(\mu T_t\|\pi) = \frac{1}{2}\operatorname{Ent}(f_t).$$

Next by Proposition 3.3 we have

$$\operatorname{Ent}(f_t) \leq e^{-4\alpha_1 t}\operatorname{Ent}(f_0) = e^{-4\alpha_1 t}D(\mu\|\pi).$$

Therefore, using the convexity of $\mu \mapsto D(\mu\|\pi)$ we have

$$\max_{\mu}\|\mu T_t - \pi\|_{\text{TV}}^2 \leq \frac{1}{2}e^{-4\alpha_1 t}\max_{\mu}D(\mu\|\pi) = \frac{1}{2}e^{-4\alpha_1 t}\log\frac{1}{\pi_{\min}}.$$

Letting the right hand side to be equal $1/(2e)^2$, we obtain the desired bound on τ_{mix}. $\square$

The reason that we express the bound of Theorem 6.2 in terms of the 2-log-Sobolev constant is that computing α_2 is usually easier than computing α_1. Nevertheless, sometimes α_1 gives a much better bound.

Let us now compare the bounds given by the above two theorems for the examples of Section 4.

For the random walk on the hypercube $\{+1,-1\}$ (Example 3) we saw that $\alpha_2 = \lambda/2 = 1/n$. Moreover, $\pi = 1/2^n$ is the uniform distribution. Then using Theorem 6.1 we obtain $\tau_{\text{mix}} = O(n^2)$ on the mixing time, while Theorem 6.2 gives $\tau_{\text{mix}} = O(n\log n)$. So the theory of log-Sobolev inequalities gives a much better bound. Indeed the advantage of the bound of Theorem 6.2, comparing to that of Theorem 6.1 is that in the former $\pi_{\min}$ appears as $\log\log(1/\pi_{\min})$ while in the latter it appears on $\log(1/\pi_{\min})$.

Another example is the random transposition (Example 4). In this case $\lambda = 2/(n-1)$ and $1/\alpha_1 = O(n)$. We also have $\pi = 1/n!$. Then Theorem 6.1 gives $\tau_{\text{mix}} = O(n^2\log n)$, while Theorem 6.2 gives $\tau_{\text{mix}} = O(n\log n)$ which was first proved in [2].

7 Analysis of boolean functions

On the applications of the hypercontractivity inequalities is in the analysis of boolean functions. Here we present an introduction to this theory. For more details we refer the reader to the survey [4].

Let $f : \{+1, -1\}^n \to \mathbb{R}$ be an arbitrary boolean function. In the following we interchangeably use two parameters $\rho \in [0, 1]$ and $t \geq 0$ that are related by

$$\rho = e^{-t}.$$

For any such ρ (or t) we define the function $T_t f : \{+1, -1\}^n \to \mathbb{R}$ as follows:

$$T_t f(y) = \sum_x \Big(\frac{1+\rho}{2}\Big)^{n-d_H(x,y)} \Big(\frac{1-\rho}{2}\Big)^{d_H(x,y)} f(x), \tag{20}$$

where $d_H(x, y)$ denotes the *Hamming distance* between $x, y \in \{+1, -1\}^n$, i.e., the number of coordinates in which x and y differ. Indeed $T_t f$ is a noisy version of f. Given $y = (b_1, \dots, b_n)$ we flip each bit b_i of y with probability $(1 - \rho)/2$, and independently of other bits, to get some random $x \in \{+1, -1\}^n$. Then we let $T_t f(y)$ to be the expectation of $f(x)$ over this random choice of x. We can state this more formally as follows.

Let (A, B) be two binary random variables jointly distributed according to

$$p(A = +1, B = +1) = p(A = -1, B = -1) = \frac{1+\rho}{4},$$

$$p(A = +1, B = -1) = p(A = -1, B = +1) = \frac{1-\rho}{4}.$$

Note that the marginal distributions of A and B are uniform. Now take $(X, Y) = (A^n, B^n)$ be n i.i.d. copies of (A, B). Then a simple calculation shows that

$$T_t f(y) = \mathbb{E}[f(X)|Y = y].$$

That is, $T_t f(y)$ is equal to the expectation of $f(X)$ conditioned on $Y = y$.

Fourier expansion: We present yet another interpretation of $T_t f$. Equip the linear space of real functions on $\{+1, -1\}^n$ with the inner product associated with the uniform distribution:

$$\langle f, g\rangle := \mathbb{E}[fg] = \frac{1}{2^n} \sum_x f(x)g(x).$$

This is a 2^n-dimensional inner product space. For any $S \subseteq [n] = \{1, \dots, n\}$ define the function $\chi_S : \{+1, -1\}^n \to \mathbb{R}$ as follows:

$$\chi_S(a_1, \dots, a_n) = \prod_{j\in S} a_j.$$

There are 2^n of these functions χ_S. Moreover, a simple calculation verifies that these functions are orthonormal to each other, i.e.,

$$\langle \chi_S, \chi_{S'}\rangle = \delta_{S,S'},$$

where $\delta_{S,S'} = 1$ if $S = S'$ and $\delta_{S,S'} = 0$ otherwise. Therefore, $\{\chi_S : S \subseteq [n]\}$ forms an orthonormal basis for the space of functions on $\{+1, -1\}^n$, and any function f can be written as a linear combination of these basis vectors:

$$f = \sum_{S \subseteq [n]} \hat{f}_S \, \chi_S.$$

The coefficients $\hat{f}_S \in \mathbb{R}$ in this expansion are called the *Fourier coefficients*, and can be computed by $\hat{f}_S = \langle f, \chi_S \rangle = \mathbb{E}[f \chi_S]$.

From definition (20) it is clear that T_t is a linear map. So to understand its action, we may first compute its action on basis vectors. In the simplest case when $n = 1$, we have two basis functions $\chi_\varnothing$ and $\chi_{\{1\}}$. Since $\chi_\varnothing$ is the constant function we have $T_t \chi_\varnothing = \chi_\varnothing$. We also have

$$T_t \chi_{\{1\}}(+1) = \frac{1+\rho}{2} - \frac{1-\rho}{2} = \rho,$$
$$T_t \chi_{\{1\}}(-1) = -\frac{1+\rho}{2} + \frac{1-\rho}{2} = -\rho.$$

Thus $T_t \chi_{\{1\}} = \rho \chi_{\{1\}}$. We leave it as an exercise for the reader to show that in general for any $S \subseteq [n]$ we have

$$T_t \chi_S = \rho^{|S|} \chi_S.$$

Thus the action of T_t on a function f is nothing but multiplying the Fourier coefficient $\hat{f}_S$ by $\rho^{|S|}$:

$$T_t f = \sum_S \hat{f}_S \rho^{|S|} \chi_S. \tag{21}$$

Now from this representation of the noise operator T_t and the fact that $\rho = e^{-t}$ it is clear that

$$T_t T_s = T_{t+s}.$$

That is, $\{T_t : t \geq 0\}$ forms a semigroup. What is more, from its definition (20) it is clear that $\{T_t : t \geq 0\}$ is indeed a Markov semigroup.

The Lindblad operator: Let us compute the generator of this semigroup. For simplicity we start with $n = 1$. In this case, we have

$$T_t \chi_\varnothing = \chi_\varnothing, \qquad \text{and} \qquad T_t \chi_{\{1\}} = e^{-t} \chi_{\{1\}}.$$

Therefore,

$$\mathcal{L} \chi_\varnothing = -\frac{\mathrm{d}}{\mathrm{d}t} T_t \chi_\varnothing \Big|_{t=0} = 0,$$

and

$$\mathcal{L} \chi_{\{1\}} = -\frac{\mathrm{d}}{\mathrm{d}t} T_t \chi_{\{1\}} \Big|_{t=0} = \chi_{\{1\}}.$$

From these two equations it is clear that $\mathcal{L} = I - \mathbb{E}$, where $\mathbb{E}$ is the expectation operator with respect to the uniform distribution.

The Lindblad operator for arbitrary n can be computed similarly. We indeed have

$$\mathcal{L} = nI - \sum_{j=1}^{n} \hat{\mathbb{E}}_j = \sum_{j=1}^{n} (I - \hat{\mathbb{E}}_j), \tag{22}$$

where $\hat{\mathbb{E}}_j$ is the lift of the expectation operator acting on the j-th coordinate, i.e., $\hat{\mathbb{E}}_j$ is the expectation operator acting on the j-th coordinate tensored with the identity operator acting on other coordinates:

$$\hat{\mathbb{E}}_j \chi_S = \begin{cases} \chi_S & j \notin S \\ 0 & \text{otherwise.} \end{cases}$$

From this we have

$$e^{t\hat{\mathbb{E}}_j} \chi_S = \begin{cases} e^t \chi_S & j \notin S \\ 1 & \text{otherwise.} \end{cases}$$

Now since the operators $\hat{\mathbb{E}}_j$ defined above commute with each other we have

$$e^{-t\mathcal{L}} \chi_S = e^{-nt} \prod_{j=1}^{n} e^{t\hat{\mathbb{E}}_j} \chi_S = e^{-nt} e^{t|S^c|} \chi_S = e^{-t|S|} \chi_S.$$

Then $e^{-t\mathcal{L}} = T_t$ and $\mathcal{L}$ is the generator of the Markov semigroup.

The Bonami-Nelson-Gross-Beckner inequality: Our next goal is to compute the log-Sobolev constants of this Markov semigroup. This can be done using the expression (22) for the Lindblad operator and Theorem 2.8. Using this theorem we have $\alpha_q(\mathcal{L}) = \alpha_q(I - \mathbb{E})$, where by $\alpha_q(I - \mathbb{E})$ we mean the q-log-Sobolev constant of the Lindblad operator for $n = 1$. We mentioned in Example 1 of Section 4 that $\alpha_2(I - \mathbb{E}) = 1/2$. As a result, using Theorem 2.4 we obtain a collection of hypercontractivity inequalities for the above Markov semigroup.

Theorem 7.1 (Bonami-Nelson-Gross-Beckner inequality)**.** *For the noise operator T_t defined in* (20) *we have*

$$\|T_t\|_{q \to p} \leq 1 \qquad \textit{iff} \qquad \sqrt{\frac{p-1}{q-1}} \leq e^t.$$

We emphasis again that based on Theorem 2.8, to prove the above theorem we only need to prove it for $n = 1$, in which case the proof reduces to establishing an inequality over a single real variable.

Theorem 7.1 is usually used for either $q = 2$ or $p = 2$ because the 2-norm can be expressed in terms of the Fourier coefficients. Using the orthonormality of χ_S's we have

$$\|f\|_2^2 = \Big\| \sum_S \hat{f}_S \chi_S \Big\|_2^2 = \sum_{S,S'} \hat{f}_S \hat{f}_{S'} \mathbb{E}[\chi_S \chi_{S'}] = \sum_S \hat{f}_S^2.$$

This is called *Parseval's identity.* Using (21) we similarly we have

$$\|T_t f\|_2 = \sum_S \rho^{2|S|} \hat{f}_S^2.$$

The following proposition states that the Fourier mass of a Boolean function with a small image is concentrated on higher degrees.

Proposition 7.2. *Let $f : \{+1, -1\}^n \to \{-1, 0, 1\}$ be a Boolean function. Then for every $\epsilon \in [0, 1]$ we have*

$$\sum_S \epsilon^{|S|} \hat{f}_S^2 \le \Pr[f(X) \neq 0]^{2/(1+\epsilon)},$$

where the probability is computed with respect to the uniform distribution on $\{+1, -1\}^n$.

Proof. Let $p = 1 + \epsilon$, $q = 2$ and $\rho = e^{-t} = \sqrt{\epsilon}$. Then using the Bonami-Nelson-Gross-Beckner inequality and Parseval's identity we have

$$\sum_S \epsilon^{|S|} \hat{f}_S^2 = \|T_t f\|_2^2 \le \|f\|_p^2.$$

Now the point is that

$$\|f\|_p^p = \mathbb{E}[|f|^p] = \Pr[f(X) \neq 0].$$

Putting these together we obtain the desired inequality. □

The degree of a Boolean function is defined by

$$\deg f = \max\{|S| : \hat{f}_S \neq 0\}.$$

By the following proposition a low-degree function does not have many *picks* because its q-norm, for every $q > 2$, is not much larger than its 2-norm.

Proposition 7.3. *Let* $\deg f = d$*. Then for every $q > 2$ we have*

$$\|f\|_q \le (q-1)^{d/2} \|f\|_2.$$

Proof. Let $\rho = e^{-t} = (q-1)^{-1/2}$ and $p = 2$. Define g by

$$g = \sum_S \rho^{-|S|} \hat{f}_S \chi_S = \sum_{S:|S| \le d} \rho^{-|S|} \hat{f}_S \chi_S.$$

Observe that $f = T_t g$, and that

$$\|g\|_2^2 = \sum_{S:|S| \le d} \rho^{-2|S|} \hat{f}_S^2 \le \sum_{S:|S| \le d} \rho^{-2d} \hat{f}_S^2 = \rho^{-2d} \|f\|_2^2.$$

Then by Theorem 7.1 we have

$$\|f\|_q^2 = \|T_t g\|_q^2 \le \|g\|_2^2 \le \rho^{-2d} \|f\|_2^2.$$

We are done. □

The following proposition states that a low-degree Boolean function is highly concentrated.

Proposition 7.4. *Let f be a Boolean function with $\mathbb{E}f = 0$ and $\mathbb{E}[f^2] = \sigma^2$. Let $d = \deg f$ and $r \geq e^{d/2}$. Then we have*

$$\Pr[|f(X)| \geq r\sigma] \leq e^{-\frac{dr^{2/d}}{2e}}.$$

Proof. Let $q = r^{2/d}/e$. By Markov's inequality

$$\Pr[|f(X)|^q \geq (r\sigma)^q] \leq \mathbb{E}[|f|^q]/(r\sigma)^q.$$

On the other hand by the previous proposition we have

$$\mathbb{E}[|f|^q]/(r\sigma)^q = \|f\|_q^q \leq (q-1)^{d/2}\|f\|_2^q.$$

Putting these together we obtain the desired inequality. □

The following theorem due to Kahn, Kalai and Linial is one of the main applications of the Bonami-Nelson-Gross-Beckner inequality in theoretical compute science. To state this theorem we need to define the *influence* of a variable. Let $f : \{+1, -1\}^n \to \{+1, -1\}$. The influence of the j-th variable is defined by

$$\mathrm{Inf}_j(f) := \Pr[f(X) \neq f(X \oplus e_j)],$$

where $X \oplus e_j$ is obtained from X by flipping its j-th coordinate. $\mathrm{Inf}_j(f)$ somehow measures the dependence of f on the j-th variable.

Let us think of such $f : \{+1, -1\}^n \to \{+1, -1\}$ as a function for a voting system in which there are n parties where the j-th party vote for $a_j \in \{+1, -1\}$. Then the outcome of the voting is $f(x) \in \{+1, -1\}$ for $x = (a_1, \ldots, a_n)$. In this case $\mathrm{Inf}_j(f)$ measures the probability, over the random choices of the votes of other parties, that the j-party can determine the outcome of the voting system. A choice of such f is the *dictator function* with $f(a_1, \ldots, a_n) = a_i$. In this case $\mathrm{Inf}_j(f)$ is equal to 1 if $j = i$ and is 0 otherwise. Another example is the *majority* function, i.e., $f(a_1, \ldots, a_n)$, say for an odd n, is equal to the majority of $a_1, \ldots, a_n$. For this function it is not hard to see that $\mathrm{Inf}_j(f) = \Theta(1/\sqrt{n})$ for all j.

One would expect that in a fair voting system the influence each party should be of order $1/n$. The question is whether such a voting systems exists or not. The following theorem excludes the existence of such a Boolean function.

Theorem 7.5. *Let $f : \{+1, -1\}^n \to \{+1, -1\}$ be a function with $\mathbb{E}f = 0$. Then there is j with $\mathrm{Inf}_j(f) \geq \log n/n$.*

The proof of this theorem, which we do not present here, is based on the ideas we developed above.

Other applications of the hypercontractivity inequality in theoretic computer science that we do not cover here are in privacy amplification, bounding the degree of approximating polynomials and inapproximability results. A good reference to learn about these results is the lecture notes of O'Donnell [5].

8 Concentration of measure inequalities

Let Z be an arbitrary random variable. Applying Markov'v inequality on the random variable $(Z - \mathbb{E}[Z])^2$ we find that

$$\Pr[|Z - \mathbb{E}Z| \geq t] = \Pr[(Z - \mathbb{E}Z)^2 \geq t^2] \leq \frac{\mathbb{E}[(Z - \mathbb{E}Z)^2]}{t^2}.$$

As a result

$$\Pr[|Z - \mathbb{E}Z| \geq t] \leq \frac{\mathrm{Var}[Z]}{t^2}.$$

This inequality is called *Chebyshev's inequality*, and is a *concentration of measure* inequality since it says that the probability that Z is far from its average is small assuming that its variance is bounded. In other words, the probability mass of Z is concentrated around its average.

Observe that if Z takes values only in the interval $[a, b]$ then $\mathrm{Var}[Z] \leq (b - a)^2/4$ so by Chebyshev's inequality we have

$$\Pr[|Z - \mathbb{E}Z| \geq t] \leq \frac{(b - a)^2}{4t^2}.$$

Our goal in this section is to derive concentration of measure inequalities that are tighter than Chebyshev's inequality. Such stronger bounds on $\Pr[|Z - \mathbb{E}Z| \geq t]$ in particular can be derived in the case where Z is of the form $Z = f(x_1, \dots, x_n)$ for some function f when x_i's are drawn independently from some distribution π on Ω.

To this end let us define

$$\psi(\theta) = \log \mathbb{E}[e^{\theta(Z - \mathbb{E}Z)}]. \tag{23}$$

$\psi(\theta)$ is called the *logarithmic moment-generation function* of Z. It can be shown as an exercise (e.g., using Hölder's inequality) that $\psi(\theta)$ is a convex function. Then again by Markov's inequality for $\theta > 0$ we have

$$\Pr[Z - \mathbb{E}[Z] \geq t] = \Pr[e^{\theta(Z - \mathbb{E}Z)} \geq e^{\theta t}] \leq e^{-(\theta t - \psi(\theta))}. \tag{24}$$

In the above inequality we may optimize over the choice of $\theta > 0$. For instance suppose that

$$\psi(\theta) \leq c\theta^2, \tag{25}$$

for some constant $c > 0$. Then we have

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-(\theta t - c\theta^2)},$$

and optimizing the right hand side of (24) over θ, i.e., letting $\theta = t/2c$, we arrive at

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-t^2/4c}.$$

The following proposition summarizes the above computations.

Lemma 8.1. *Let Z be an arbitrary random variable and define $\psi(\theta)$ as in* (23)*. Suppose that for some constant $c > 0$ we have $\psi(\theta) \le c\theta^2$ for all $\theta > 0$. Then for all $t \ge 0$ we have*

$$\Pr[Z - \mathbb{E}[Z] \ge t] \le e^{-t^2/4c}. \tag{26}$$

We note that if Z is a Gaussian random variable with zero mean and variance σ^2 then $\psi(\theta) = \sigma^2\theta^2/2$. Thus Lemma 8.1 can in particular be applied when Z is Gaussian. Because of this, a random variable Z satisfying (26) is called to have *Gaussian concentration.*

Hoeffding's inequality: Let us give an example for illustrating a quadratic upper bounds on $\psi(\theta)$. Assume that $X_1, \dots, X_n$ are i.i.d. random variables with zero-mean $\mathbb{E}[X_i] = 0$ and $a \le X_i \le b$. Let $Z = (X_1 + \cdots + X_n)/\sqrt{n}$. Our goal is to prove a quadratic upper bound on $\psi(\theta)$ for this choice of Z. We compute

$$\mathbb{E}\Big[e^{\theta Z}\Big] = \mathbb{E}\Big[\prod_i e^{\frac{\theta}{\sqrt{n}}X_i}\Big] = \mathbb{E}\Big[e^{\frac{\theta}{\sqrt{n}}X_1}\Big]^n. \tag{27}$$

Now using the convexity of the exponential function we have

$$e^{\frac{\theta}{\sqrt{n}}X_1} \le \frac{b - X_1}{b-a} e^{\frac{\theta}{\sqrt{n}}a} + \frac{X_1 - a}{b-a} e^{\frac{\theta}{\sqrt{n}}b}.$$

Taking expectation we find that

$$\mathbb{E}\big[e^{\frac{\theta}{\sqrt{n}}X_1}\big] \le \frac{b}{b-a} e^{\frac{\theta}{\sqrt{n}}a} + \frac{-a}{b-a} e^{\frac{\theta}{\sqrt{n}}b} \le e^{\frac{(b-a)^2}{8n}\theta^2}. \tag{28}$$

We now need the following lemma whose proof can be found in Appendix B.

Lemma 8.2. *For any $p \in [0,1]$ and $x \in \mathbb{R}$ we have*

$$(1-p)e^{-px} + pe^{(1-p)x} \le e^{\frac{x^2}{8}}.$$

In the above lemma let $p = -a/(b-a)$ and $x = \theta(b-a)/\sqrt{n}$. Note that since $a \le X_1 \le b$ and $\mathbb{E}[X_1] = 0$ we have $a \le 0 \le b$, and then $p \in [0,1]$. Therefore, (28) gives

$$\mathbb{E}\big[e^{\frac{\theta}{\sqrt{n}}X_1}\big] \le e^{\frac{(b-a)^2}{8n}\theta^2}.$$

Putting all these together we arrive at[1]

$$\psi(\theta) \le \frac{(b-a)^2}{8}\theta^2.$$

Thus by Lemma 8.1 we obtain the Hoeffding's inequality

$$\Pr\Big[\frac{X_1 + \cdots + X_n}{\sqrt{n}} \ge t\Big] \le e^{-\frac{2}{(b-a)^2}t^2}. \tag{29}$$

We used the special form of $Z = (X_1 + \cdots + X_n)/\sqrt{n}$ in the first step (27). Nevertheless, what we really need is a *martingale type property* there. A similar inequality as above can be proven for martingales that is called *Azuma's inequality.*

[1] This inequality is called Hoeffding's lemma.

General case: Let us consider the more general case where $Z = f(X_1, \ldots, X_n)$ is an arbitrary function of i.i.d. X_i's distributed according to π. For simplicity also assume that $\mathbb{E}Z = 0$. The idea is to use log-Sobolev inequalities for the function $e^{\theta Z}$ to prove a quadratic upper bound on $\psi(\theta)$ defined in (23). This is called *Herbst's argument.*

We have

$$\operatorname{Ent}(e^{\theta Z}) = \theta \mathbb{E}[Ze^{\theta Z}] - \mathbb{E}[e^{\theta Z}] \log \mathbb{E}[e^{\theta Z}] = \theta\big(e^{\psi(\theta)}\big)' - e^{\psi(\theta)}\psi(\theta).$$

Therefore,

$$\frac{\operatorname{Ent}(e^{\theta Z})}{\theta^2 e^{\psi(\theta)}} = \Big(\frac{\psi(\theta)}{\theta}\Big)'.$$

As a result if we show that $\operatorname{Ent}(e^{\theta Z}) \leq c\theta^2 e^{\psi(\theta)}$, we conclude that

$$\Big(\frac{\psi(\theta)}{\theta}\Big)' \leq c,$$

and then by integration we obtain the desired inequality $\psi(\theta) \leq c\theta^2$. As a summary, the problem of proving Gaussian concentration reduces to the problem of proving upper bounds of form $c\theta^2 e^{\psi(\theta)}$ on $\operatorname{Ent}(e^{\theta Z})$. This is where log-Sobolev inequalities enter.

We use the 1-log-Sobolev inequality for the Dirichlet form

$$\mathcal{L} = \sum_{i=1}^{n} (I - \hat{\mathbb{E}}_i)$$

where $\mathbb{E}_i$ is the expectation with respect to the i-th coordinate (with distribution π). Let $\alpha_1 = \alpha_1(\mathcal{L}) = \alpha_1(I - \mathbb{E})$. Then letting $Z_i = f(X_1, \ldots X_i', \ldots X_n)$ where X_i' is an independent copy of X_i, we have

$$4\alpha_1 \operatorname{Ent}(e^{\theta Z}) \leq \mathcal{E}(\theta Z, e^{\theta Z}) = \frac{1}{2} \sum_i \mathbb{E}\Big[(\theta Z - \theta Z_i)\big(e^{\theta Z} - e^{\theta Z_i}\big)\Big],$$

where the equality follows from (11). Now as an exercise one can verify that for every $a, b \in \mathbb{R}$ we have

$$(b-a)(e^b - e^a) \leq \frac{1}{2}(b-a)^2(e^a + e^b).$$

Therefore,

$$4\alpha_1 \operatorname{Ent}(e^{\theta Z}) \leq \frac{1}{4}\theta^2 \sum_i \mathbb{E}\Big[(Z - Z_i)^2\big(e^{\theta Z} + e^{\theta Z_i}\big)\Big] = \frac{1}{2}\theta^2 \sum_i \mathbb{E}\Big[(Z - Z_i)^2 e^{\theta Z}\Big]. \tag{30}$$

Let us define

$$\gamma = \max_{x, x^{(1)}, \ldots, x^{(n)}} \sum_i \big(f(x) - f(x^{(i)})\big)^2,$$

where the maximum is over all $x = (x_1, \ldots, x_n)$ and $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ with the condition that $x_j^{(i)} = x_j$ for all $i \neq j$. Hence,

$$\operatorname{Ent}(e^{\theta f}) \leq \frac{\gamma}{8\alpha_1}\theta^2 e^{\psi(\theta)}.$$

With the previous arguments we conclude that

$$\psi(\theta) \leq \frac{\gamma}{8\alpha_1}\theta^2,$$

and then by Lemma 8.1

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-2\alpha_1 t^2/\gamma}. \tag{31}$$

We note that $\alpha_2 \leq \alpha_1$. Therefore, the above argument also shows that

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-2\alpha_2 t^2/\gamma}.$$

□

The above Gaussian concentration inequality, although useful in most applications, is not as tight as Hoeffding's inequality. Indeed, if we let $Z = (X_1 + \cdots + X_n)/\sqrt{n}$, where X_i's are i.i.d. with $\mathbb{E}[X_i] = 0$ and $a \leq X_i \leq b$, then $\gamma \leq (b-a)^2$. On the other hand, we know that $\alpha_1(I - \mathbb{E}) \geq 1/4$ for any distribution π. Putting these in (31) we arrive at

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-\frac{t^2}{2(b-a)^2}}.$$

Here, comparing to (29), the exponent of the right hand side is not tight.

In the following we will use Herbst's argument for proving a concentration of measure inequality that is tighter than (31), and does imply Hoeffding's inequality. Before working on the general case, it is instructive to first consider the spacial case where X_i's are *binary* random variables with *uniform* distribution, i.e., π is the uniform distribution over the binary set $\{+1, -1\}$. The point is that in this case we have $\alpha_1(I - \mathbb{E}) \geq \alpha_2(I - \mathbb{E}) = 1/2$. Moreover, in this case, Z and Z_i as defined above, are equal with probability $1/2$. Therefore, (30) can be rewritten as

$$2\,\mathrm{Ent}(e^{\theta Z}) \leq \frac{1}{4}\theta^2 \sum_i \mathbb{E}\Big[(Z - Z_i')^2 e^{\theta Z}\Big] \leq \frac{1}{4}\gamma\theta^2 e^{\psi(\theta)},$$

where $Z_i' = f(X_1, \ldots, X_{i-1}, -X_i, X_{i+1}, \ldots, X_n)$. This gives $\psi(\theta) \leq \frac{1}{8}\gamma\theta^2$ and then

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-\frac{2t^2}{\gamma}}.$$

We see in the above example of the uniform distribution over a binary set that a more careful analysis gives a tighter concentration of measure inequality. In the following to generalize this approach for arbitrary distributions, we use another inequality that resembles 1-log-Sobolev inequalities [16] .

Theorem 8.3. *Let $X_1, \ldots, X_n$ be arbitrary independent random variables and let $Z = f(X_1, \ldots, X_n)$. Let $V_i = g_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ be an arbitrary function of $X_1, \ldots, X_n$ except the i-th one. Then we have*

$$\mathrm{Ent}(e^Z) \leq \sum_{i=1}^{n} \mathbb{E}\big[\ell(Z - V_i)e^Z\big],$$

where $\ell(x) = e^{-x} + x - 1$.

Proof. Using the subadditivity of the entropy function (Theorem 2.7) it suffices to prove the theorem in the base case $n = 1$, i.e., for any random variable Z and any constant c we have

$$\mathrm{Ent}(e^Z) \leq \mathbb{E}[\ell(Z-c)e^Z].$$

This is a simple calculus exercise. Define

$$h(c) = \mathbb{E}\big[\ell(Z-c)e^Z\big] = e^c + \mathbb{E}\big[Ze^Z\big] - c\mathbb{E}\big[e^Z\big] - \mathbb{E}\big[e^Z\big],$$

Observe that $h(c)$ is a convex function which is minimized at $c = \ln \mathbb{E}[e^Z]$. Then for every c we have

$$\mathbb{E}[\ell(Z-c)e^Z] \geq h\big(\ln \mathbb{E}[e^Z]\big) = \mathrm{Ent}(e^Z).$$

□

We are now ready to complete Herbst's argument and prove *McDiarmid's inequality.*

Theorem 8.4. *Let $X_1, \ldots, X_n$ be independent random variables. Let $Z = f(X_1, \ldots, X_n)$ where f is an arbitrary function and define*

$$\gamma = \max_{x, x^{(1)}, \ldots, x^{(n)}} \sum_i \big(f(x) - f(x^{(i)}\big)^2,$$

where the maximum is over all $x = (x_1, \ldots, x_n)$ and $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ with the condition that $x_j^{(i)} = x_j$ for all $i \neq j$. Then we have

$$\psi(\theta) \leq \frac{\gamma}{8}\theta^2,$$

where as before $\psi(\theta)$ is the logarithmic moment generation function given by (23)*. As a result we have*

$$\Pr[Z - \mathbb{E}[Z] \geq t] \leq e^{-\frac{2t^2}{\gamma}}.$$

Proof. For simplicity assume that $\mathbb{E}[Z] = 0$. Applying Herbst's argument, it suffices to show that for every θ we have $\mathrm{Ent}(e^{\theta Z}) \leq \frac{\gamma}{8}\theta^2 e^{\psi(\theta)}$. To this end we use the inequality of Theorem 8.3. For arbitrary random variables $V_1, \ldots, V_n$ where V_i does not depend on X_i, we have

$$\mathrm{Ent}(e^{\theta Z}) \leq \sum_{i=1}^n \mathbb{E}\big[\ell\big(\theta Z - V_i\big)e^{\theta Z}\big],$$

where $\ell(t) = e^{-t} + t - 1$. Define the random variables

$$A_i = \inf_{x_i} f(X_1, \ldots, x_i, \ldots X_n), \qquad B_i = \sup_{x_i} f(X_1, \ldots, x_i, \ldots X_n).$$

Since $\ell(t)$ is a convex function we have

$$\ell(\theta Z - V_i) \leq \max\big\{\ell(\theta A_i - V_i), \ell(\theta B_i - V_i)\big\}.$$

On the other hand, letting

$$V_i = \theta A_i + \ln(\theta C_i) - \ln(1 - e^{-\theta C_i}),$$

where $C_i = B_i - A_i$ we have

$$\ell(\theta A_i - V_i) = \ell(\theta B_i - V_i) = \frac{\theta C_i}{1 - e^{-\theta C_i}} + \ln \frac{1 - e^{-\theta C_i}}{\theta C_i} - 1.$$

We conclude that

$$\mathrm{Ent}(e^{\theta Z}) \leq \sum_{i=1}^{n} \mathbb{E}\big[q(\theta C_i) e^{\theta Z}\big], \tag{32}$$

where

$$q(t) = \frac{t}{1 - e^{-t}} + \ln \frac{1 - e^{-t}}{t} - 1.$$

We claim that for any t we have $q(t) \leq t^2/8$. To verify this recall that in Lemma 8.2 we showed that

$$\ln\Big((1-p)e^{-pt} + pe^{(1-p)t}\Big) \leq t^2/8, \qquad \forall p \in [0,1].$$

Optimizing this inequality over the choice of p and letting

$$p = \frac{e^t - t - 1}{t(e^t - 1)} \in [0,1],$$

we find that $q(t) \leq t^2/8$. Using this in (32) we obtain

$$\mathrm{Ent}(e^{\theta Z}) \leq \frac{1}{8}\theta^2 \mathbb{E}\Big[\Big(\sum_{i=1}^{n} C_i^2\Big) e^{\theta Z}\Big].$$

Now observe that by assumption $\sum_{i=1}^{n} C_i^2 \leq \gamma$. Therefore,

$$\mathrm{Ent}(e^{\theta Z}) \leq \frac{\gamma}{8}\theta^2 \mathbb{E}\big[e^{\theta Z}\big],$$

as desired.

□

The above arguments are also valid in the continuous case.

Theorem 8.5. *Let $f : \mathbb{R}^k \to \mathbb{R}$ be a 1-Lipschitz function, i.e., for all $\mathbf{x} \in \mathbb{R}^k$ we have $|\nabla f(\mathbf{x})| \leq 1$. Then we have*

$$\Pr[f - \mathbb{E}[f] \geq t] \leq e^{-\frac{t^2}{2}},$$

where the probability and expectation are with respect to the k-dimensional standard normal distribution with density (16).

Proof. For simplicity assume that $\mathbb{E}[f] = 0$. Applying Herbst's argument, we need to bound $\mathrm{Ent}(e^{\theta f})$. Using the 1-log-Sobolev inequality (15) and its tensorization we have

$$\mathrm{Ent}(e^{\theta f}) \leq \frac{\theta^2}{2}\mathbb{E}\big[|\nabla f|^2 e^{\theta f}\big] \leq \frac{\theta^2}{2}\mathbb{E}\big[e^{\theta f}\big],$$

where the second inequality follows from the assumption that f is 1-Lipschitz and then $|\nabla f|^2 \leq 1$. The above inequality gives the desired result.

□

One can prove the above theorem using the central limit theorem and Theorem 8.4.

9 Transportation-cost inequalities

Concentration of measure inequalities are closely related to *transportation-cost inequalities.* In this section we explain the notion of transportation-cost inequalities and their connections to log-Sobolev inequalities. For a detailed study of the subject we refer to [14].

Suppose that (Ω, d, π) is a metric probability space. For any two distributions μ, ν on Ω, and $p \geq 1$ we define the *p-Wasserstein* distance between μ, ν by

$$W_p(\mu, \nu) := \inf_{\xi} \Big(\sum_{x,y\in\Omega} \xi(x,y) d^p(x,y) \Big)^{1/p}, \tag{33}$$

where infimum is over all distribution ξ on $\Omega \times \Omega$ whose marginals are μ and ν, i.e., over all *couplings* of μ, ν. It can be shown that $W_p(\cdot,\cdot)$ is indeed a metric on the space of distributions, and that

$$W_p(\mu,\nu) \leq W_q(\mu,\nu), \qquad \forall 1 \leq p \leq q. \tag{34}$$

The problem of computing the Wasserstein distance $W_p^p(\mu,\nu)$ is an optimization of a linear function on ξ with linear constraints, i.e., it is a linear program. Thus using the *strong duality* of linear programs one can express $W_p^p(\mu,\nu)$ as a maximization problem. The result would be the following theorem called *Kantorovich duality.*

Theorem 9.1. *For every two measure μ, ν and $p \geq 1$ we have*

$$W_p^p(\mu,\nu) = \sup\big\{\mathbb{E}_\mu[g] + \mathbb{E}_\nu[h] : \ \ g(x) + h(y) \leq d^p(x,y)\ \forall x, y\big\}. \tag{35}$$

In the case of $p = 1$ we can further simplify the above formula. The following theorem is called the *Kantorovich-Rubinstein theorem.*

Theorem 9.2. *For every μ, ν we have*

$$W_1(\mu,\nu) = \sup_{f:\text{1-Lipschitz}} \big|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]\big|,$$

where supremum is taken over all 1-Lipschitz functions $f : \Omega \to \mathbb{R}$.

Proof. By Theorem 9.1 we have

$$W_1(\mu,\nu) = \sup\big\{\mathbb{E}_\mu[g] + \mathbb{E}_\nu[h] : \; g(x) + h(y) \leq d(x,y) \ \forall x, y\big\}. \tag{36}$$

Take optimal functions g, h in the above optimization. Define

$$f(x) := \inf_y d(x,y) - h(y).$$

Since $g(x) + h(y) \leq d(x,y)$ we have $f(x) \geq g(x)$. Moreover, $f(x) \leq d(x,x) - h(x) \leq -h(x)$. Then we have

$$\mathbb{E}_\mu[g] + \mathbb{E}_\nu[h] \leq \mathbb{E}_\mu[f] - \mathbb{E}_\nu[f].$$

It is also easy to show that f is 1-Lipschitz. Therefore,

$$W_1(\mu,\nu) \leq \sup_{f:\text{1-Lipschitz}} \big|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]\big|.$$

Inequality in the other direction follows once in (36) we put $g = -h = f$ in which case $g(x) + h(y) \leq d(x,y)$ would reduce to f being 1-Lipschitz.

□

Definition 9.3. A metric probability space (Ω, d, π) is called to satisfy the *p-transportation-cost inequality* with constant $c > 0$ (also called the *p-Talagrand inequality*) denoted by $T_p(c)$ if for every distribution μ we have

$$W_p(\mu,\pi) \leq \sqrt{2cD(\mu\|\pi)},$$

where as before, $D(\mu\|\pi)$ is the KL-divergence (7).

Observe that because of (34), $T_q(c)$ implies $T_p(c)$ if $1 \leq p \leq q$.

Example: Consider the metrix $d(x,y) = 1 - \delta_{x,y}$. Then as an exercise one can verify that

$$W_1(\mu,\nu) = \|\mu - \nu\|_{\text{TV}}.$$

Then, by Pinsker's inequality (19) we have

$$W_1(\mu,\pi) \leq \sqrt{\frac{1}{2}D(\mu\|\pi)}.$$

This means that for any π the space (Ω, d, π) satisfies $T_1(1/4)$. □

A crucial observation in the study of transportation-cost inequalities is the *Donsker-Varadhan formula:*

$$D(\mu\|\pi) = \sup_f \mathbb{E}_\mu[f] - \ln \mathbb{E}_\pi[e^f],$$

where the supremum is over all functions $f : \Omega \to \mathbb{R}$. This formula essentially says that the *convex conjugate* or *Legendre transform* of the convex function $\mu \mapsto D(\mu\|\pi)$ is equal to the convex function $f \mapsto \ln \mathbb{E}_\pi[e^f]$. Since applying the Legendre transform on a nice function twice, we get the starting function itself, we also have

$$\ln \mathbb{E}_\pi[e^f] = \sup_\mu \mathbb{E}_\mu[f] - D(\mu\|\pi). \tag{37}$$

Now that according to Theorem 9.1 and the Donsker-Varadhan formula we have variational expressions for both sides of a transportation-cost inequality, we can derive their dual equivalent version. The following two theorems are due to Bobkov and Götze [6].

Theorem 9.4. *(Ω, d, π) satisfies $T_1(c)$ if and only if for every 1-Lipschitz function $f : \Omega \to \mathbb{R}$ with $\mathbb{E}_\pi[f] = 0$ we have*

$$\psi(\theta) \leq \frac{c}{2}\theta^2,$$

where $\psi(\theta) = \log \mathbb{E}[e^{\theta f}]$. In particular, $T_1(c)$ implies that

$$\Pr[f - \mathbb{E}[f] \geq t] \leq e^{-\frac{t^2}{2c}},$$

for all 1-Lipschitz functions f.

Proof. By the Kantorovich-Rubinstein formula (Theorem 9.2) transportation-cost inequality $T_1(c)$ holds iff

$$\mathbb{E}_\mu[f] - \mathbb{E}_\pi[f] \leq \sqrt{2cD(\mu\|\pi)},$$

for all 1-Lipschitz functions f. Moreover, we have

$$\sqrt{2cD(\mu\|\pi)} \leq \frac{c\theta}{2} + \frac{D(\mu\|\pi)}{\theta}, \qquad \forall \theta > 0,$$

and equality holds for some $\theta > 0$. Therefore, $T_1(c)$ holds iff

$$\mathbb{E}_\mu[f] - \mathbb{E}_\pi[f] \leq \frac{c\theta}{2} + \frac{D(\mu\|\pi)}{\theta},$$

for all 1-Lipschitz f and $\theta > 0$. Since this inequality holds for all probability distributions μ we find that $T_1(c)$ is equivalent to

$$\sup_\mu \mathbb{E}_\mu[\theta f] - D(\mu\|\pi) \leq \frac{c\theta^2}{2} + \theta\mathbb{E}_\pi[f]$$

Using the dual of the Donsker-Varadhan formula (37) we find that $T_1(c)$ is equivalent to

$$\ln \mathbb{E}_\pi[e^{\theta f}] \leq \frac{c\theta^2}{2} + \theta\mathbb{E}_\pi[f],$$

being satisfied for all 1-Lipschitz f and $\theta > 0$.

□

By this theorem the connection between log-Sobolev inequalities and transportation-cost inequalities will be clear. We saw previously that log-Sobolev inequalities can be used to prove quadratic upper bounds on the logarithmic moment generation functions. The above theorem says that such bounds give transportation-cost inequalities, so log-Sobolev inequalities imply transportation-cost inequalities. In particular, the above theorem, together with Theorem 8.4 imply Pinsker's inequality.

Corollary 9.5. *The metric probability space* $(\Omega^n, d_{1,n}, \pi^n)$ *with*

$$d_{1,n}(x^n, y^n) = \sum_{i=1}^{n}(1 - \delta_{x_i,y_i})$$

satisfies $T_1(n/4)$*.*

Proof. In Theorem 8.4 we showed that for every function f on Ω^n we have

$$\psi(\theta) \le \frac{\gamma}{8}\theta^2,$$

with

$$\gamma = \max_{x,x^{(1)},\ldots,x^{(n)}} \sum_i (f(x) - f(x^{(i)})^2.$$

where the maximum is over all $x = (x_1, \ldots, x_n)$ and $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ with the condition that $x_j^{(i)} = x_j$ for all $i \ne j$. Now the point is that if $f : \Omega^n \to \mathbb{R}$ is 1-Lipschitz (with respect to the above metric) then $\gamma \le n$. Then the desired result follows from the previous theorem. □

The constant of the transportation-cost inequality given by the above corollary scales with n. Nevertheless, in certain settings one can derive transportation-cost (and then concentration of measure) inequalities that do not depend on n. To explore this let us first compute a dual formulation for T_2 inequalities.

Theorem 9.6. *Let* (Ω, d, π) *be a metric probability space. Then the followings are equivalent:*

(i) (Ω, d, π) *satisfies* $T_2(c)$ *inequality for some* $c > 0$*.*

(ii) For all functions g, h *with* $g(x) + h(y) \le d^2(x, y)$ *we have*

$$\ln \mathbb{E}_\pi[e^{\frac{1}{2c}h}] \le -\frac{1}{2c}\mathbb{E}_\pi[g].$$

Proof. The proof is similar to that of Theorem 9.4. By the Kantorovich duality, (i) is equivalent to

$$\mathbb{E}_\mu[h] + \mathbb{E}_\pi[g] \le 2cD(\mu\|\pi),$$

for all distributions μ and functions g, h such that $g(x)+h(y) \le d^2(x, y), \forall x, y$. Optimizing over μ and using (37) we obtain the equivalent formulation (ii).

□

The dual formulations of T_1 and T_2 inequalities given by Theorem 9.4 and Theorem 9.6 can easily be extended to all $1 \le p \le 2$.

Theorem 9.7. *Let (Ω, d, π) be a metric probability space and $1 < p < 2$. Then the followings are equivalent:*

(i) (Ω, d, π) satisfies $T_p(c)$ inequality for some $c > 0$.

(ii) For all functions g, h with $g(x) + h(y) \le d^p(x, y)$ we have

$$\ln \mathbb{E}_\pi[e^{\frac{t}{pc}h}] \le -\frac{t}{pc}\mathbb{E}_\pi[g] + \frac{2-p}{2pc}t^{\frac{2}{2-p}}, \qquad \forall t \ge 0.$$

Proof. Use Young's inequality

$$a^{\frac{2}{p}} = \sup_{t\ge 0} \frac{2}{p}at - \frac{2-p}{p}t^{\frac{2}{2-p}},$$

and follow similar steps as in the proofs of Theorem 9.4 and Theorem 9.6. □

Similar to log-Sobolev inequalities, transportation-cost inequalities also enjoy a tensorization property. The following theorem can be proven directly using properties of p-norms or using the dual formulations of T_p inequalities given by the above theorems. We leave its proof as an exercise for the reader.

Theorem 9.8. *Suppose that (Ω, d, π) satisfies $T_p(c)$ for some $1 \le p \le 2$. Then for every n the space $(\Omega^n, d_{p,n}, \pi^n)$ satisfies $T_p(cn^{2/p-1})$ where*

$$d_{p,n}(x^n, y^n) := \Big(\sum_{i=1}^{n} d^p(x_i, y_i)\Big)^{1/p}. \tag{38}$$

The above theorem, in particular says that $T_2(c)$ for space (Ω, d, π) gives $T_2(c)$ for $(\Omega^n, d_{2,n}, \pi^n)$. That is, T_2 inequalities are *dimension-free*, so one may obtain *dimenstion-free* concentration of measure inequalities out of them.

Theorem 9.9. *[7] Let (Ω, d, π) be a metric probability space. Then the followings are equivalent:*

(i) (Ω, d, π) satisfies $T_2(c)$ inequality for some $c > 0$.

(ii) For all n and all functions $f : \Omega^n \to n$ that are 1-Lipschitz with respect to the norm $d_{2,n}$ defined in (38) we have $\psi(\theta) \le c\theta^2/2$ and then

$$\Pr\left[f - \mathbb{E}_\pi[f] \ge t\right] \le e^{-\frac{t^2}{2c}}.$$

Going from (i) to (ii) is easy: first use the tensorization property of T_2 inequalities, and then the fact that 1-Wasserstein distance is dominated by 2-Wasserstein distance.

Finally use Theorem 9.4. Proof of the other direction is more involved for which we refer to [7].

Now the question is how we can prove T_2 transportation-cost inequalities. In the following we will show that in certain situations 1-log-Sobolev inequalities imply T_2 inequalities and then dimension-free concentration of measure inequalities. The following theorem is due to Otto and Villani.

Theorem 9.10. *Let π be a probability measure on $\mathbb{R}^k$. Suppose that π satisfies the following 1-log-Sobolev inequality*

$$4\alpha_1 \operatorname{Ent}_\pi(e^f) \le \int_{\mathbb{R}^k} |\nabla f|^2 e^f \mathrm{d}\pi. \tag{39}$$

Then $(\mathbb{R}^k, d, \pi)$, where $d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \Big(\sum_i (x_i - y_i)^2\Big)^{1/2}$ is the Euclidean norm, satisfies $T_2(1/2\alpha_1)$.

We called (39) a 1-log-Sobolev inequality since it is associated with the Lindblad operator $\mathcal{L} = \nabla V \cdot \nabla - \Delta$ that is reversible with respect to π when $\mathrm{d}\pi = e^{-V}\mathrm{d}\mathbf{x}$ (see (17)). As an exercise one can verify that the 1-log-Sobolev inequality associated with the Lindblad operator has the form (39).

Proof. We will present two proofs for this theorem.

First proof (sketch): We know that 1-log-Sobolev inequalities imply Gaussian concentration of measure inequalities. On the other hand, 1-log-Sobolev inequalities have the tensorization property. Thus, π^n as a distribution on $\mathbb{R}^{nk}$ satisfies concentration of measure inequality with respect to the Euclidean norm, that is dimension-free. Then the desired result follows from Theorem 9.9.

Second proof: The above proof is based on Theorem 9.9 whose proof is non-trivial and is not presented in this manuscript. So we present a self-contained proof here.

For every $f : \mathbb{R}^k \to \mathbb{R}$ we define

$$Q_t f(\mathbf{x}) = \inf_{\mathbf{y}\in\mathbb{R}^k} f(\mathbf{y}) + \frac{1}{t}|\mathbf{x} - \mathbf{y}|^2, t > 0, \qquad Q_0 f = f. \tag{40}$$

It is well-known that $(t, \mathbf{x}) \mapsto Q_t f(\mathbf{x})$ is a solution of the Hamilton-Jacobi equation:

$$\begin{cases} \frac{\partial}{\partial t} u + \frac{1}{4}|\nabla_{\mathbf{x}} u|^2 = 0, & t \ge 0, \mathbf{x} \in \mathbb{R}^k \\ u(0, \mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^k. \end{cases} \tag{41}$$

The formula (40) as a solution of the above differential equation is called the Hopf-Lax formula and is proven in Appendix C. We also show there that $\{Q_t : t \ge 0\}$ forms a semigroup.

Now for a fixed $t \ge 0$, we apply the given 1-log-Sobolev inequality to the function $\mathbf{x} \mapsto \alpha_1 t Q_t f(\mathbf{x})$:

$$\operatorname{Ent}_\pi\big(e^{\alpha_1 t Q_t f}\big) \le \frac{1}{4\alpha_1}\int_{\mathbb{R}^k} |\nabla_{\mathbf{x}} \alpha_1 t Q_t f(\mathbf{x})|^2 e^{\alpha_1 t Q_t f(\mathbf{x})} \mathrm{d}\pi = -\alpha_1 t^2 \int_{\mathbb{R}^k} \Big(\frac{\partial}{\partial t} Q_t f(\mathbf{x})\Big) e^{\alpha_1 t Q_t f(\mathbf{x})} \mathrm{d}\pi, \tag{42}$$

where equality follows from (41). Define

$$\varphi(t) := \mathbb{E}\Big[e^{\alpha_1 t Q_t f}\Big] = \int_{\mathbb{R}^k} e^{\alpha_1 t Q_t f(\mathbf{x})} \mathrm{d}\pi.$$

Using (42) we have

$$\begin{aligned} t\varphi'(t) &= \mathbb{E}\Big[\alpha_1 t Q_t f e^{\alpha_1 t Q_t f}\Big] + \alpha_1 t^2 \mathbb{E}\Big[\Big(\frac{\partial}{\partial t} Q_t f(\mathbf{x})\Big) e^{\alpha_1 t Q_t f}\Big] \\ &= \operatorname{Ent}\big(e^{\alpha_1 t Q_t f}\big) + \varphi(t)\ln\varphi(t) + \alpha_1 t^2 \mathbb{E}\Big[\Big(\frac{\partial}{\partial t} Q_t f(\mathbf{x})\Big) e^{\alpha_1 t Q_t f}\Big] \\ &\le \varphi(t)\ln\varphi(t). \end{aligned}$$

This implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\ln\varphi(t)}{t}\Big) \le 0,$$

and then

$$\ln \mathbb{E}\Big[e^{\alpha_1 Q_1 f}\Big] = \ln\varphi(1) \le \lim_{t\to 0^+} \frac{\ln\varphi(t)}{t} = \lim_{t\to 0^+} \frac{\varphi'(t)}{\varphi(t)} = \alpha_1 \mathbb{E}[f]. \tag{43}$$

Now in order to prove $T_2(1/2\alpha_1)$ we use its equivalent characterization given by Theorem 9.6. Let g, h be two functions satisfying $g(\mathbf{x}) + h(\mathbf{y}) \le |\mathbf{x}-\mathbf{y}|^2$. Then

$$h(\mathbf{y}) \le \inf_{\mathbf{x}} -g(\mathbf{x}) + |\mathbf{y}-\mathbf{x}|^2 = Q_1(-g)(\mathbf{y}).$$

As a result, using (43) for $f = -g$ we have

$$\ln \mathbb{E}\Big[e^{\alpha_1 h}\Big] \le \ln \mathbb{E}\Big[e^{\alpha_1 Q_1(-g)}\Big] \le -\alpha_1 \mathbb{E}[g].$$

This gives the desired result. □

The above theorem has been generalized for other metric spaces, particularly to discrete ones. We refer the reader to [8, 9, 10, 11, 13] and references therein for such generalizations.

Since we already know that $\alpha_1 \ge 1/2$ for the standard normal distribution we obtain the following.

Corollary 9.11. *(Talagrand's inequality) The Euclidian space $\mathbb{R}^k$ with the multi-dimensional standard normal distribution satisfies $T_2(1)$.*

We also point out here that the second proof of Theorem 9.10 gives the following hypercontractivity inequality.

Theorem 9.12. *[17] Let π be a probability measure on $\mathbb{R}^k$. Then the 1-log-Sobolev inequality*

$$4\alpha_1 \operatorname{Ent}_\pi(e^f) \le \int_{\mathbb{R}^k} |\nabla f|^2 e^f \mathrm{d}\pi, \tag{44}$$

holds for all function f if and only if for any $p \ge 0$ and any function f we have

$$\big\|e^{Q_t f}\big\|_{p+t\alpha_1} \le \|e^f\|_p, \qquad \forall t \ge 0$$

where $Q_t f$ is defined in (40).

Observe that inequality (43) that we showed in the proof of Theorem 9.10 is the special case of the above theorem for $t = 1$ and $p = 0$.

Proof. Define

$$G(t) = \ln \|e^{Q_t f}\|_{p(t)},$$

with $p(t) = p + \alpha_1 t$. Then we have

$$\frac{1}{\alpha_1} p(t)^2 e^{G(t)} G'(t) = \operatorname{Ent}_\pi\big(e^{p(t)Q_t f}\big) - \frac{1}{4\alpha_1}\mathbb{E}\big[|\nabla\big(p(t)Q_t f\big)|^2 e^{p(t)Q_t f}\big] \leq 0.$$

Then if the 1-log-Sobolev inequality (44) holds, we have $G'(t) \leq 0$ for all $t \geq 0$. Therefore, for every $t \geq 0$ we have

$$G(t) \leq G(0) = \ln \|e^f\|_p.$$

Conversely, letting $p = 1$, if $G(t) \leq G(0)$ for all $t \geq 0$, we have $G'(0) \leq 0$ which is equivalent to the desired log-Sobolev inequality.

□

10 Isoperimetric inequalities

Isoperimetric inequalities are inequalities that bound the surface area of a set from below in terms of its volume. Here we start with Harper's edge isoperimetric inequality in the hypercube, and explain its connection to log-Sobolev inequalities.

Theorem 10.1. *Let $A \subseteq \{0,1\}^n$ be arbitrary and define $E(A, A^c)$ to be the set of edges from A to its complement in the Boolean hypercube. Then we have*

$$|E(A, A^c)| \geq |A| \cdot (n - \log_2 |A|).$$

Moreover, equality holds if A is a sub-hypercube.

We do not have a proof of this theorem using log-Sobolev inequalities. Nevertheless, loosing a factor of $\ln 2$, a weaker version of this isoperimetric inequality easily follows from log-Sobolev inequalities.[2]

Proof of a waker version. Let f be the indicator function of the set A. Consider the Lindblad operator $\mathcal{L} = \sum_i (I - \hat{\mathbb{E}}_i)$ as before in (22) in which expectations are with respect to the uniform distribution. Recall that $\alpha_2(\mathcal{L}) = \alpha_2(I - \mathbb{E}) = 1/2$. Therefore,

$$\frac{1}{2}\operatorname{Ent}(f^2) \leq \mathbb{E}[f\mathcal{L}f] = \frac{1}{2}\sum_i \mathbb{E}\big[\big(f(X) - f(X_i)\big)^2\big],$$

[2] A proof of this theorem can be derived by a simple induction on n and using the inequality

$$(x+y)\log_2(x+y) \geq x\log_2 x + y\log_2 y + 2\min\{x, y\}.$$

where X_i is obtained from X by randomly changing its i-th coordinate. Since f and then f^2 takes values in $\{0,1\}$ we have

$$\mathrm{Ent}(f^2) = -\frac{|A|}{2^n}\ln\frac{|A|}{2^n},$$

and

$$\sum_i \mathbb{E}\big[\big(f(X)-f(X_i)\big)^2\big] = \frac{1}{2^n}|E(A,A^c)|.$$

Putting these together we arrive at

$$|E(A,A^c)| \geq \ln 2\cdot|A|(n-\log_2|A|).$$

□

Let us now state another inequality, known as the *blowing-up lemma*. This lemma was first proved by Margulis [25]. Here we present a simpler proof due to Marton [26].

Theorem 10.2 (Blowing-up lemma). *Let Ω be an arbitrary finite set, and let $A\subseteq\Omega^n$ be a subset. Define*

$$A_r = \{x\in\Omega^n : d_H(x,A)\leq r\},$$

where as before $d_H(x,y)=\sum_{i=1}^n(1-\delta_{x_i,y_i})$ is the Hamming distance and $d_H(x,A)=\min_{y\in A}d_H(x,y)$. Suppose that

$$r\geq(1+\epsilon)\sqrt{\frac{n}{2}\log\frac{|\Omega|^n}{|A|}},$$

where $\epsilon>0$. Then we have

$$|A_r|\geq|\Omega|^n\Big(1-e^{-\epsilon^2\log(|\Omega^n|-|A|)}\Big).$$

To understand the content of this theorem let us assume that $|A|=c|\Omega|^n$ for some constant $c>0$, i.e., the size of $A\subseteq\Omega^n$ is a constant fraction of $|\Omega^n|$. Then the theorem says that as long as $r\geq(1+\epsilon)\sqrt{(n\log 1/c)/2}$ we have

$$|A_r|\geq|\Omega^n|(1-e^{\epsilon^2\log 1/c}).$$

In other words, although $|A|$ could be a constant fraction of $|\Omega^n|$, the size of its r-neighborhood for $r=O(\sqrt{n})$, is almost equal to the size Ω^n.

Proof. Let us equip the space Ω with the uniform distribution denoted by π. As mentioned before, by Pinsker's inequality, (Ω,d,π) with $d(x,y)=1-\delta_{x,y}$ satisfies $T_1(1/4)$. Then using the tensorization of Talagrand's inequalities (Theorem 9.8) or Corollary 9.5 the space (Ω^n,d_H,π^n) satisfies $T_1(n/4)$. That is for any distribution μ on Ω^n we have

$$W_1(\mu,\pi^n)\leq\sqrt{\frac{n}{2}D(\mu\|\pi^n)}.$$

Let us assume that $\mu = \mu_A$ is the conditional measure supported on A defined by

$$\mu_A(x) = \begin{cases} \frac{1}{|A|}, & x \in A \\ 0. & \text{otherwise.} \end{cases}$$

Then a simple computation verifies that

$$D(\mu_A \| \pi^n) = \log \frac{1}{\pi^n(A)} = \log \frac{|\Omega^n|}{|A|}.$$

Putting in the Talagrand's inequality we obtain

$$W_1(\mu_A, \pi^n) \leq \sqrt{\frac{n}{2} \log \frac{1}{\pi^n(A)}}.$$

Using the triangle inequality and the above inequality twice we have

$$W_1(\mu_A, \mu_{A_r^c}) \leq W_1(\mu_A, \pi^n) + W_1(\mu_{A_r^c}, \pi^n) \leq \sqrt{\frac{n}{2} \log \frac{1}{\pi^n(A)}} + \sqrt{\frac{n}{2} \log \frac{1}{\pi^n(A_r^c)}},$$

where $A_r^c = \Omega^n \setminus A$ is the complement of A_r.

The next step is to give a lower bound on $W_1(\mu_A, \mu_{A_r^c})$. Suppose that $\xi(x, y)$ is a coupling of μ_A and $\mu_{A_r^c}$. Since μ_A is supported on A only and the marginal of ξ on the first coordinate equals μ_A we have $\xi(x, y) = 0$ for all $x \notin A$. Similarly we have $\xi(x, y) = 0$ for all $y \notin A_r^c$. Therefore,

$$\sum_{x,y \in \Omega^n} \xi(x, y) d_H(x, y) = \sum_{x \in A, y \in A_r^c} \xi(x, y) d_H(x, y) \geq r,$$

where in the last step we use the fact that for all $(x, y) \in A \times A_r^c$ we have $d_H(x, y) \geq r$.

$$r \leq \sqrt{\frac{n}{2} \log \frac{1}{\pi^n(A)}} + \sqrt{\frac{n}{2} \log \frac{1}{\pi^n(A_r^c)}}.$$

Rearranging this inequality and using $\pi^n(A_r^c) = 1 - \pi^n(A_r)$ we obtain the desired result. $\square$

Observe that as the above proof shows, a similar inequality can be stated for any distribution π and not just the uniform distribution.

We now turn to the Gaussian isoperimetric inequality in $\mathbb{R}^k$. Recall that the standard normal distribution on $\mathbb{R}^k$ has the density

$$\mathrm{d}\pi(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{|\mathbf{x}|^k}{2}} \mathrm{d}\mathbf{x}.$$

That is the volume of a Borel set A with respect to this measure is given by

$$\pi(A) = \frac{1}{(2\pi)^{k/2}} \int_A e^{-\frac{|\mathbf{x}|^k}{2}} \mathrm{d}\mathbf{x}.$$

We also define the r-Euclidean open neighborhood of A by

$$A_r := \{\mathbf{x} \in \mathbb{R}^k : \, |\mathbf{x} - \mathbf{A}| < \mathbf{r}\},$$

where $|\mathbf{x} - A| = \inf_{\mathbf{y} \in y} |\mathbf{x} - \mathbf{y}|$ is the Euclidean distance of $\mathbf{x}$ from A. The *Gaussian surface measure* of a Borel set A is defined by

$$\pi_s(\partial A) := \liminf_{r \to 0^+} \frac{\pi(A_r) - \pi(A)}{r}$$

To state the main theorem let

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad \Phi(x) = \int_{-\infty}^{x} \varphi(y) \mathrm{d}y,$$

be the density and cumulative distribution function of the one-dimensional standard normal distribution.

Theorem 10.3. *For any Borel set $A \subseteq \mathbb{R}^k$ we have*

$$\pi_s(\partial A) \geq \varphi \circ \Phi^{-1}(\pi(A)), \tag{45}$$

and equality holds if A is a half-space.

Let us first examine the equality case. Fix some vector $\mathbf{u} \in \mathbb{R}^k$ with $|\mathbf{u}| = 1$ define the half-space

$$H_a = \{\mathbf{x} \in \mathbb{R}^k : \, \langle \mathbf{x}, \mathbf{u} \rangle < a\},$$

where $\langle \mathbf{x}, \mathbf{u} \rangle$ is the Euclidean inner product. Using the rotation invariance of Gaussian distributions, we see that $\pi(H_a)$ is independent of the choice of $\mathbf{u}$, and for $\mathbf{u} = (1, 0, \cdots, 0)$ it is not hard to verify that $\pi(H_a) = \Phi(a)$ and

$$\pi_s(\partial H_a) = \lim_{r \to 0^+} \frac{\pi(H_{a+r}) - \pi(H_a)}{r} = \Phi'(a) = \varphi(a).$$

Therefore, (45) turns to an equality for H_a. In other words, half-spaces have minimal Gaussian surface areas.

A functional version of Theorem 10.3 has been derived by Bobkov [15].

Theorem 10.4. *For every Lipschitz function $f : \mathbb{R}^k \to [0, 1]$ we have*

$$\Psi\Big(\int_{\mathbb{R}^k} f \mathrm{d}\pi\Big) \leq \int_{\mathbb{R}^k} \sqrt{\Psi^2(f) + |\nabla f|^2} \, \mathrm{d}\pi, \tag{46}$$

where $\Psi = \varphi \circ \Phi^{-1}$.

The above theorem easily gives (45). The idea is to let f to be the characteristic function of the set A. Note however that the the characteristic function is not Lipschitz, so in (46) we pick

$$f_r(\mathbf{x}) := \max\Big\{1 - \frac{1}{r}|\mathbf{x} - A|, 0\Big\},$$

and then take the limit of $r \to 0^+$. Assuming that A is a nice set, f_r tends to the characteristic function of A, and then $\Psi \circ f_r$ tends to 0 since $\Psi(0) = \Psi(1) = 0$. We then obtain

$$\Psi(\pi(A)) \leq \liminf_{r \to 0^+} \int_{\mathbb{R}^k} |\nabla f_r| \mathrm{d}\pi.$$

On the other hand, f_r is constant on A and on $\mathbb{R}^k \setminus \overline{A_r}$. Thus ∇f_r is zero on A and on $\mathbb{R}^k \setminus \overline{A_r}$. Moreover, from the definition we have $|\nabla f_r| \leq 1/r$ on $\overline{A_r} \setminus A$. Therefore,

$$\Psi(\pi(A)) \leq \liminf_{r \to 0^+} \int_{\mathbb{R}^k} |\nabla f_r| \mathrm{d}\pi \leq \liminf_{r \to 0^+} \frac{\pi(A_r) - \pi(A)}{r} = \pi_s(\partial A).$$

□

We now turn to the the proof of Theorem 10.4. The original proof of Bobkov [15] of this theorem is by first stating a discrete version of it, and then based on the central limit theorem, generalizing it to the Gaussian case (similarly to the second proof of Theorem 5.5). Here we given another proof, which although is not based on log-Sobolev or hypercontractivity inequalities, is based on the properties of the Ornstein-Uhlenbeck semigroup.

Proof of Theorem 10.4. Let $f : \mathbb{R}^k \to [0, 1]$ be sufficiently smooth. Let T_t denote the Ornstein-Uhlenbeck semigroup given by (12). Define

$$J(t) := \int_{\mathbb{R}^k} \sqrt{\Psi^2(T_t f) + |\nabla T_t f|^2} \, \mathrm{d}\pi.$$

$J(0)$ equal the right hand side of (46). Moreover, since $T_t f$ tends to the constant function $m = \int_{\mathbb{R}^k} f \mathrm{d}\pi$ as $t \to \infty$, we have

$$\lim_{t \to \infty} J(t) = \int_{\mathbb{R}^k} \Psi(m) \mathrm{d}\pi.$$

Then the desired inequality is equivalent to $J(\infty) \leq J(0)$. For this it suffices to show that $J(t)$ is a non-increasing function, i.e., $J'(t) \leq 0$. In the following, for simplicity of presentation we assume that $k = 1$, and prove $J'(t) \leq 0$. The proof for arbitrary k is identical.

Recall that the generator of the Ornstein-Uhlenbeck semigroup is $\mathcal{L}g(x) = xg'(x) - g''(x)$. For simplicity of notation let $Kf(x) = \Psi^2(f(x)) + |f'(x)|^2$ and $T_t f = f_t$. We then have

$$J'(t) = \int_{\mathbb{R}} \frac{1}{\sqrt{K f_t(x)}} \Big(- \mathcal{L} f_t(x) \cdot \Psi'(f_t(x)) \cdot \Psi(f_t(x)) - f_t'(x) \cdot \mathcal{L} f_t'(x) \Big) \mathrm{d}\pi$$

Now recall that in Lemma 5.3 we showed that

$$\int_{\mathbb{R}} g(x) \cdot \mathcal{L}h(x) \mathrm{d}\pi = \int_{\mathbb{R}} g'(x) h'(x) \mathrm{d}\pi.$$

Using this and the facts that

$$(Kf_t)'(x) = 2f_t'(x) \cdot \Psi'(f_t(x)) \cdot \Psi(f_t(x)) + 2f_t'(x) \cdot f_t''(x),$$

and $\Psi(y) \cdot \Psi''(y) = -1$ which can be proven as an exercise, we find that

$$J'(t) = -\int_{\mathbb{R}} \frac{1}{Kf_t(x)^{3/2}} \Big(\Psi'(f_t(x))f_t'^2 - \Psi(f_t(x))f_t''(x)\Big)^2 \mathrm{d}\pi \leq 0.$$

We are done.

□

We present yet another connection between log-Sobolev inequalities and isoperimetric inequalities. We show that how the Brunn-Minkowski inequality gives a proof of log-Sobolev inequality associated to Gaussian measures and in fact to all measures satisfying the Bakry-Emery criterion (see Theorem 5.7).

The Brunn-Minkowski inequality states that for any two compact subsets $A, B \subset \mathbb{R}^k$ we have

$$\mathrm{Vol}(A)^{1/k} + \mathrm{Vol}(B)^{1/k} \leq \mathrm{Vol}(A + B)^{1/k},$$

where $\mathrm{Vol}(A)$ denotes the volume of the set A with respect to the Lebesgue measure: $\mathrm{Vol}(A) = \int_A \mathrm{d}\mathbf{x}$, and $A + B$ is the Minkowski sum:

$$A + B = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in A, \mathbf{y} \in Y\}.$$

Before continuing let us derive an isoperimetric inequality using the Brunn-Minkowski inequality. Let $B_\epsilon = B_\epsilon(0)$ be the ball of radius $\epsilon > 0$ around the origin. We have $\mathrm{Vol}(B_\epsilon) = \epsilon^k \mathrm{Vol}(B_1)$. Moreover, $A_\epsilon = A + B_\epsilon$ is the *ϵ-neighborhood* of A. Then the Brunn-Minkowski inequality gives

$$\mathrm{Vol}(A)^{1/k} + \epsilon \mathrm{Vol}(B_1)^{1/k} \leq \mathrm{Vol}(A_\epsilon)^{1/k},$$

and equivalently

$$\frac{1}{\epsilon}\big(\mathrm{Vol}(A_\epsilon)^{1/k} - \mathrm{Vol}(A)^{1/k}\big) \geq \mathrm{Vol}(B_1)^{1/k}.$$

Taking the limit $\epsilon \to 0^+$ we obtain

$$\mathrm{Area}(\partial A) \geq k \mathrm{Vol}(B_1)^{1/k}\, \mathrm{Vol}(A)^{1-1/k}, \tag{47}$$

where

$$\mathrm{Area}(\partial A) = \lim_{\epsilon \to 0^+} \frac{\mathrm{Vol}(A_\epsilon) - \mathrm{Vol}(A)}{\epsilon},$$

is the *surface area* of A. Thus (47) bounded the surface area of a set in terms of its volume and is an isoperimetric inequality.

Now let us turn in to the proof of the Brunn-Minkowski inequality. We first state and prove a *functional* version of the Brunn-Minkowski inequality called the Prékopa-Leindler inequality.

Theorem 10.5. *Let $u, v, w : \mathbb{R}^k \to [0, +\infty)$ be measurable functions such that for some $\theta \in (0, 1)$ we have*

$$w(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \geq u(\mathbf{x})^\theta v(\mathbf{y})^{1-\theta}, \quad \forall \mathbf{x}, \mathbf{y}.$$

Then we have

$$\int_{\mathbb{R}^k} w(\mathbf{x})\mathrm{d}\mathbf{x} \geq \Big(\int_{\mathbb{R}^k} u(\mathbf{x})\mathrm{d}\mathbf{x}\Big)^\theta \Big(\int_{\mathbb{R}^k} v(\mathbf{x})\mathrm{d}\mathbf{x}\Big)^{1-\theta}.$$

By a change of variable the above theorem can be written as follows. Suppose that

$$w(\mathbf{x} + \mathbf{y}) \geq u(\mathbf{x})^\theta v(\mathbf{y})^{1-\theta}, \quad \forall \mathbf{x}, \mathbf{y}. \tag{48}$$

Then we have

$$\int_{\mathbb{R}^k} w(\mathbf{x})\mathrm{d}\mathbf{x} \geq \frac{1}{\theta^{k\theta}(1-\theta)^{k(1-\theta)}} \Big(\int_{\mathbb{R}^k} u(\mathbf{x})\mathrm{d}\mathbf{x}\Big)^\theta \Big(\int_{\mathbb{R}^k} v(\mathbf{x})\mathrm{d}\mathbf{x}\Big)^{1-\theta}. \tag{49}$$

Let w, u, v be the characteristic functions of the set $A+B, A, B$ respectively. Then (48) holds for all $\theta \in (0, 1)$ from the definitions. Then for all $\theta \in (0, 1)$ we have we find that

$$\mathrm{Vol}(A+B) \geq \frac{1}{\theta^{k\theta}(1-\theta)^{k(1-\theta)}} \mathrm{Vol}(A)^\theta \mathrm{Vol}(B)^{1-\theta}.$$

Optimizing over the choice θ and letting

$$\theta = \frac{\mathrm{Vol}(A)^{1/k}}{\mathrm{Vol}(A)^{1/k} + \mathrm{Vol}(B)^{1/k}},$$

the Brunn-Minkowski inequality is obtained.[3]

Proof of Theorem 10.5. We prove the equivalent formulation of the theorem given by equations (48) and (49). We sketch a proof by induction on k. For $k = 1$ we note that for any non-negative function f

$$\int_{\mathbb{R}} f(x)\mathrm{d}x = \int_0^\infty \mathrm{Vol}\big(L_f(t)\big)\mathrm{d}t,$$

where $L_f(t) = f^{-1}(t, \infty)$. Then (48) says that

$$L_w(t) \supseteq L_u(t) + L_v(t), \qquad \forall t \in \mathbb{R}.$$

Therefore, by the 1-dimensional Brunn-Minkowski inequality, whose proof is easy, we have

$$\mathrm{Vol}(L_w(t)) \geq \mathrm{Vol}(L_u(t) + L_v(t)) \geq \mathrm{Vol}(L_u(t)) + \mathrm{Vol}(L_v(t)).$$

Integrating over t we obtain

$$\int_{\mathbb{R}} w(x)\mathrm{d}x \geq \int_{\mathbb{R}} u(x)\mathrm{d}x + \int_{\mathbb{R}} v(x)\mathrm{d}x \geq \frac{1}{\theta^\theta(1-\theta)^{1-\theta}} \Big(\int_{\mathbb{R}} u(x)\mathrm{d}x\Big)^\theta \Big(\int_{\mathbb{R}} v(x)\mathrm{d}x\Big)^{1-\theta},$$

[3] Another proof of the Brunn-Minkowski inequality can be derived from the convexity of certain functions over the space of probability measures equipped with the 2-Wasserstein distance.

where for the second inequality we use $\theta a + (1-\theta)b \geq a^\theta + b^{1-\theta}$.

For the induction step we use a tensorization type argument. For any $x, y, z \in \mathbb{R}$ define $u_x, v_y, w_z : \mathbb{R}^{k-1} \to \mathbb{R}$ by

$$u_x(\mathbf{x}) := u(x, \mathbf{x}), \quad v_y(\mathbf{y}) := u(y, \mathbf{y}), \quad w_z(\mathbf{z}) := u(z, \mathbf{z}).$$

These functions satisfy the assumption (48). Then by the induction hypothesis we have

$$\int_{\mathbb{R}^{k-1}} w_{x+y}(\mathbf{x})\mathrm{d}\mathbf{x} \geq \frac{1}{\theta^{(k-1)\theta}(1-\theta)^{(k-1)(1-\theta)}} \Big(\int_{\mathbb{R}^{k-1}} u_x(\mathbf{x})\mathrm{d}\mathbf{x} \Big)^\theta \Big(\int_{\mathbb{R}^{k-1}} v_y(\mathbf{x})\mathrm{d}\mathbf{x} \Big)^{1-\theta}.$$

This means that the functions

$$\tilde{w}(z) = \theta^{(k-1)\theta}(1-\theta)^{(k-1)(1-\theta)} \int_{\mathbb{R}^{k-1}} w_z(\mathbf{x})\mathrm{d},$$

and

$$\tilde{u}(x) = \int_{\mathbb{R}^{k-1}} u_x(\mathbf{x})\mathrm{d}\mathbf{x}, \qquad \tilde{v}(y) = \int_{\mathbb{R}^{k-1}} v_y(\mathbf{x})\mathrm{d}\mathbf{x},$$

defined on $\mathbb{R}$ satisfy (48). Then the desired result follows from the theorem for $k = 1$. $\square$

We now prove a log-Sobolev inequality from the Brunn-Minkowski inequality and its functional version from [17].

Proof of Theorem 5.7. For a function $g : \mathbb{R}^k \to \mathbb{R}$ and $t > 0$ define

$$u(\mathbf{x}) = e^{\frac{1}{\theta}Q_t g(\mathbf{x}) - V(\mathbf{x})}, \qquad v(\mathbf{x}) = e^{-V(\mathbf{x})}, \qquad w(\mathbf{x}) = e^{g(\mathbf{x}) - V(\mathbf{x})},$$

where $\theta = 2/(2 + ct) \in (0, 1)$ and $Q_t g$ is given by (40). From the definitions for every $\mathbf{x}, \mathbf{y}$ we have

$$\begin{aligned} Q_t g(\mathbf{x}) &\leq g(\theta\mathbf{x} + (1-\theta)\mathbf{y}) + \frac{1}{t}\big|\mathbf{x} - (\theta\mathbf{x} + (1-\theta)\mathbf{y})\big|^2 \\ &= g(\theta\mathbf{x} + (1-\theta)\mathbf{y}) + \frac{c\theta(1-\theta)}{2}|\mathbf{x} - \mathbf{y}|^2. \end{aligned}$$

On the other hand, since by assumption $\mathrm{Hess}(V) \geq cI$ we have

$$\theta V(\mathbf{x}) + (1-\theta)V(\mathbf{y}) - V(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \geq \frac{c\theta(1-\theta)}{2}|\mathbf{x} - \mathbf{y}|^2.$$

Putting these together we find that

$$Q_t g(\mathbf{x}) - \theta V(\mathbf{x}) - (1-\theta)V(\mathbf{y}) \leq g(\theta\mathbf{x} + (1-\theta)\mathbf{y}) - V(\theta\mathbf{x} + (1-\theta)\mathbf{y}),$$

which is equivalent to

$$u(\mathbf{x})^\theta v(\mathbf{y})^{1-\theta} \leq w(\theta\mathbf{x} + (1-\theta)\mathbf{y}).$$

They by Theorem 10.5 we have

$$\int_{\mathbb{R}^k} w(\mathbf{x})\mathrm{d}\mathbf{x} \geq \Big(\int_{\mathbb{R}^k} u(\mathbf{x})\mathrm{d}\mathbf{x} \Big)^\theta \Big(\int_{\mathbb{R}^k} v(\mathbf{x})\mathrm{d}\mathbf{x} \Big)^{1-\theta},$$

which can be written as

$$\mathbb{E}_\pi[e^g] \geq \mathbb{E}_\pi[e^{\frac{1}{\theta}Q_t g}]^\theta.$$

This means that

$$\|e^g\|_1 \geq \|e^{Q_t g}\|_{1+\frac{c}{2}t}, \qquad \forall t \geq 0.$$

Then by Theorem 9.12 we have

$$2c\,\mathrm{Ent}_\pi(e^g) \leq \int_{\mathbb{R}^k} |\nabla g|^2 e^g \mathrm{d}\pi.$$

Letting $g = 2\ln f$ we obtain the desired log-Sobolev inequality.

□

A related inequality to isoperimetric inequalities is the *Loomis-Whitney inequality* which in its special form states that for every measurable set A in $\mathbb{R}^k$ we have

$$\mathrm{Vol}(A) \leq \prod_{i=1}^k \mathrm{Vol}(A_i)^{1/(k-1)},$$

where A_i is the $(k-1)$-dimensional projection of A in the direction of the i-th coordinate, one and $\mathrm{Vol}(A_i)$ is the $(k-1)$-dimensional volume of A_i. The Loomis-Whiney inequality itself is a special case of the Brascamp-Lieb inequality.

11 Markov semigroups as gradient flows

In this section we show that the flow generated by a Markov semigroup on the space of probability measures can be thought of as a gradient flow when the space of probability measures is equipped with the 2-Wasserstein distance. This idea is originated in the seminal work of Otto [22].

In this section we restrict ourself to the Ornstein-Uhlenbeck semigroup associated to $\mathrm{d}\pi = e^{-V}\mathrm{d}\mathbf{x}$, i.e., $\mathcal{L} = \nabla V \cdot \nabla - \Delta$. Recall that the space of real functions on $\mathbb{R}^k$ is equipped with the inner product

$$\langle f, g\rangle_\pi = \int_{\mathbb{R}^k} f(\mathbf{x})g(\mathbf{x})\mathrm{d}\pi(\mathbf{x}).$$

Moreover, $\mathcal{L}$ is reversible with respect to π, i.e.,

$$\langle f, \mathcal{L}g\rangle_\pi = \langle \mathcal{L}f, g\rangle_\pi = \langle \nabla f, \nabla g\rangle_\pi = \int_{\mathbb{R}^k} \nabla f \cdot \nabla g \,\mathrm{d}\pi.$$

We will also use the notation $\langle\cdot,\cdot\rangle$ with any subscript to denote the inner product with respect to the Lebesgue measure:

$$\langle f, g\rangle = \int_{\mathbb{R}^k} f(\mathbf{x})g(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{50}$$

We will also frequently use the fact that div and $-\nabla$ are adjoints of each other with respect to this inner product, i.e., using integration by parts we have

$$\langle f, \mathrm{div}(g) \rangle = \int_{\mathbb{R}^k} f \mathrm{div}(g) \mathrm{d}\mathbf{x} = -\int_{\mathbb{R}^k} \nabla f \cdot g \, \mathrm{d}\mathbf{x} = -\langle \nabla f, g \rangle. \tag{51}$$

Finally for simplicity of notation we use $\partial_t = \partial/\partial t$ and for $1 \leq i \leq k$ we denote $\partial_i = \partial/\partial x_i$

Let μ_t be a sufficiently smooth curve in the space of probability measures on $\mathbb{R}^k$ that are absolutely continuous with respect to π and let

$$\mathrm{d}\mu_t = f_t \mathrm{d}\pi. \tag{52}$$

Since μ_t's are probability measures, the integral of $\partial_t \mu_t$ over the whole space is zero. Thus $\partial \mu_t$ can be written as the divergence of some function: $\partial_t \mu_t = -\mathrm{div}(a_t)$. Next we may write $a_t = \mu_t b_t$ and $\partial_t \mu_t = -\mathrm{div}(\mu_t b_t)$. This equation does not uniquely determines b_t. However, we may choose the *shortest* such function b_t with respect to the inner product

$$\langle b_t, b_t \rangle_{\mu_t} = \int_{\mathbb{R}^k} |b_t|^2 \mathrm{d}\mu_t.$$

It is not hard to verify that such a shortest vector must be orthogonal, with respect to the above inner product, to all function c_t with $\mathrm{div}(\mu_t c_t) = 0$. Thus, using (51), such a b_t can itself be written as the gradient of another function $b_t = \nabla \psi_t$:

$$\partial_t \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t).$$

With this equation in mind, we may equip the space of probability measures with a Riemannian structure. We say that the tangent vector to the above curve at point μ_t is $\nabla \psi_t$ with norm

$$\langle \nabla \psi_t, \nabla \psi_t \rangle_{\mu_t} = \int_{\mathbb{R}^k} |\nabla \psi_t|^2 \mathrm{d}\mu_t. \tag{53}$$

Benamou-Brenier formula: Let $\{\mu_t : 0 \leq t \leq 1\}$ be an arbitrary curve in the space of probability measures determined by

$$\partial \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t).$$

Then the length of this curve, under the Riemannian metric (53) equals

$$\int_0^1 \langle \nabla \psi_t, \nabla \psi_t \rangle_{\mu_t}^{1/2} \mathrm{d}t = \int_0^1 \Big(\int_{\mathbb{R}^k} |\nabla \psi_t|^2 \mathrm{d}\mu_t \Big)^{1/2} \mathrm{d}t.$$

Moreover, the geodesic distance between two probability measures μ_0, μ_1 is given by

$$\inf_{\{\mu_t : 0 \leq t \leq 1\}} \left\{ \int_0^1 \Big(\int_{\mathbb{R}^k} |\nabla \psi_t|^2 \mathrm{d}\mu_t \Big)^{1/2} \mathrm{d}t \, : \, \partial_t \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t) \right\}.$$

It is well-known that the above equation is equal to the squared of 2-Wasserstein distance between μ_0, μ_1 with respect to the Euclidean metric and is called the *Benamou-Brenier formula*:

$$W_2(\mu_0, \mu_1) = \inf_{\{\mu_t : 0 \le t \le 1\}} \left\{ \int_0^1 \Big(\int_{\mathbb{R}^k} |\nabla \psi_t|^2 \mathrm{d}\mu_t \Big)^{1/2} \mathrm{d}t \, : \, \partial_t \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t) \right\}.$$

Moreover, since geodesics have constant speed[4] we also have

$$W_2^2(\mu_0, \mu_1) = \inf_{\{\mu_t : 0 \le t \le 1\}} \left\{ \int_0^1 \int_{\mathbb{R}^k} |\nabla \psi_t|^2 \mathrm{d}\mu_t \mathrm{d}t \, : \, \partial_t \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t) \right\}.$$

Thus the metric induced by the Riemannian structure (53) on the space of probability measures is the 2-Wasserstein distance. In Appendix D we derive the geodesic equations in this metric.

□

Let F be the entropy function relative to π, i.e.,

$$F(\mu_t) = D(\mu_t \| \pi) = \mathrm{Ent}_\pi(f_t) = \int_{\mathbb{R}^k} f_t \log f_t \mathrm{d}\pi,$$

where f_t is given by (52). We claim that the *gradient flow* of the function $F(\cdot)$ on the space of probability measures equipped with the above Riemannian metric is given by the Ornstein-Uhlenbeck semigroup. To prove over claim, recall that the curve $\{\mu_t : t \ge 0\}$ is the gradient flow of F if the tangent vectors to this curve are given by the minus of the gradients of F. In other words,

$$\partial_t \mu_t = -\mathrm{div}\big(\mu_t(-\mathrm{grad} F(\mu_t))\big) = \mathrm{div}\big(\mu_t \, \mathrm{grad} F(\mu_t)\big). \tag{54}$$

Here we denote the gradient of F with notation $\mathrm{grad}F$ to highlight the fact that it is computed with respect to the aforementioned Riemannian structure. That is, $\mathrm{grad}(F)$ is given by

$$\langle \mathrm{grad} F(\mu_t), \nabla \psi_t \rangle_{\mu_t} = \partial_t F(\mu_t). \tag{55}$$

Let us first compute $\mathrm{grad}F(\mu_t)$:

$$\begin{aligned}
\partial_t F(\mu_t) &= \int_{\mathbb{R}^k} \partial_t (f_t \log f_t) \mathrm{d}\pi \\
&= \int_{\mathbb{R}^k} \partial_t f_t \log f_t \mathrm{d}\pi \\
&= \int_{\mathbb{R}^k} \partial \mu_t \log f_t \mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^k} \mathrm{div}(\mu_t \nabla \psi_t) \log f_t \mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^k} \mu_t \nabla \psi_t \cdot \nabla(\log f_t) \mathrm{d}\mathbf{x} \\
&= \langle \nabla \psi_t, \nabla(\log f_t) \rangle_{\mu_t}
\end{aligned}$$

[4] Consider a re-parametrization of a geodesic.

Then comparing to (55) we conclude that

$$\mathrm{grad} F(\mu_t) = \nabla(\log f_t).$$

We now use this in the gradient flow equation (54):

$$\begin{aligned}\partial_t \mu_t &= \mathrm{div}(\mu_t\, \mathrm{grad} F(\mu_t))\\ &= \mathrm{div}(\mu_t\, \nabla(\log f_t))\\ &= \nabla\mu_t \cdot \nabla(\log f_t) + \mu_t \Delta(\log f_t).\end{aligned}$$

Therefore,

$$\begin{aligned}\partial_t f_t = \frac{1}{\pi}\partial_t \mu_t &= \frac{1}{\pi}\Big(\nabla(\pi f_t)\cdot\nabla(\log f_t) + \pi f_t \Delta(\log f_t)\Big)\\ &= \Big(\frac{\nabla \pi}{\pi} f_t + \nabla f_t\Big)\cdot \nabla(\log f_t) + f_t \Delta(\log f_t)\\ &= (-V f_t + \nabla f_t)\cdot \frac{\nabla f_t}{f_t} + f_t \frac{f_t \Delta f_t - |\nabla f_t|^2}{f_t^2}\\ &= -V\nabla f_t + \Delta f_t\\ &= -\mathcal{L} f_t.\end{aligned}$$

We conclude that the gradient flow of $F(\cdot)$, i.e., the entropy function, is nothing but the flow given by the generator $\mathcal{L} = \nabla V \cdot \nabla - \Delta$ of the Ornstein-Uhlenbeck semigroup. □

With this geometric picture of the flows of Markov semigroups we can now give new proofs and generalize some of the previous results. Let us start with a proof of the Otto-Villani theorem. We first need a lemma.

Lemma 11.1. *Let $f_t = e^{-t\mathcal{L}} f_0$ with $\mathcal{L}$ be the generator of the Ornstein-Uhlenbeck semigroup with respect to $\pi = e^{-V}$, such that $\mu_t = f_t \pi$ is a probability measure. Then for any probability measure τ we have*

$$\partial_t W_2(\mu_t, \tau) \le \sqrt{\langle f_t, \mathcal{L} \log f_t\rangle_\pi}.$$

Proof. By the triangle inequality we have

$$\partial_t W_2(\mu_t,\tau) = \lim_{s\to 0^+} \frac{1}{s}\Big(W_2(\mu_{t+s},\tau) - W_2(\mu_t,\tau)\Big) \le \lim_{s\to 0^+}\frac{1}{s} W_2(\mu_t, \mu_{t+s}).$$

Next letting $\nabla\psi_t$ be the tangent vector to μ_t, i.e., $\partial_t \mu_t = -\mathrm{div}(\mu_t \nabla\psi_t)$, by the Benamou-Brenier formula we have

$$W_2(\mu_t, \mu_{t+s}) \le \int_t^{t+s} \langle \nabla\psi_v, \nabla\psi_v\rangle_{\mu_v}^{1/2} \mathrm{d}v.$$

On the other hand, since $\{\mu_t : t \ge 0\}$ is a gradient flow of the entropy function we have

$$\partial_t\, \mathrm{Ent}_\pi(f_t) = \langle \nabla\psi_t, \mathrm{grad}\ \mathrm{Ent}_\pi(f_t)\rangle_{\mu_t} = -\langle \nabla\psi_t, \nabla\psi_t\rangle_{\mu_t}.$$

Therefore,

$$W_2(\mu_t, \mu_{t+s}) \le \int_t^{t+s} \sqrt{-\partial_v \operatorname{Ent}_\pi(f_v)}\mathrm{d}v = \int_t^{t+s} \sqrt{\langle f_v, \mathcal{L} \log f_v\rangle_\pi}\mathrm{d}v.$$

Taking the limit we find that

$$\begin{aligned}
\partial_t W_2(\mu_t, \tau) &\le \lim_{s\to 0^+} \frac{1}{s} W_2(\mu_t, \mu_{t+s}) \\
&\le \lim_{s\to 0^+} \frac{1}{s} \int_t^{t+s} \sqrt{\langle f_v, \mathcal{L} \log f_v\rangle_\pi}\mathrm{d}v \\
&= \sqrt{\langle f_t, \mathcal{L} \log f_t\rangle_\pi}.
\end{aligned}$$

□

Now we can present yet another proof of the result of Otto and Villani.

Proof of Theorem 9.10. Let $f_t = e^{-t\mathcal{L}} f_0$ and μ_t be as in the previous lemma. Define

$$F(t) = W_2(\mu_t, \mu_0) + \sqrt{\frac{1}{\alpha_1} \operatorname{Ent}_\pi(f_t)}.$$

We compute

$$\begin{aligned}
\partial_t F(t) &= \partial_t W_2(\mu_t, \mu_0) + \frac{\partial_t \operatorname{Ent}_\pi(f_t)}{2\sqrt{\alpha_1 \operatorname{Ent}_\pi(f_t)}} \\
&\le \sqrt{\langle f_t, \log f_t\rangle_\pi} - \frac{\langle f_t, \mathcal{L} \log f_t\rangle_\pi}{2\sqrt{\alpha_1 \operatorname{Ent}_\pi(f_t)}} \\
&\le 0,
\end{aligned}$$

where in the first inequality we use Lemma 11.1 and in the second inequality we use the 1-log-Sobolev inequality $4\alpha_1 \operatorname{Ent}_\pi(f_t) \le \langle f_t, \log f_t\rangle_\pi$ in the assumption. Therefore, $F(t)$ is non-increasing function and we have

$$\sqrt{\frac{1}{\alpha_1} \operatorname{Ent}_\pi(f_0)} = F(0) \ge \lim_{t\to\infty} F(t) = W_2(\pi, \mu_0),$$

where in computing the limit we use the fact that $\mu_t \to \pi$ and $f_t \to 1$ as t tends to infinity. We are done as the choice of f_0 is arbitrary.

□

12 HWI inequality

In this section we prove the so called HWI inequality [23]. This inequality involves the three quantities of interest, namely, the entropy, the derivative of entropy which is the Dirichlet form, and the 2-Wasserstein distance.

Theorem 12.1 (HWI inequality). *Let π be a Borel probability measure on $\mathbb{R}^k$ with $\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x$ where $\mathrm{Hess}(V) \geq cI$ for some constant c. Then for every Borel probability measure μ with $\mu = f\pi$ we have*

$$D(\mu\|\pi) = \mathrm{Ent}_\pi(f) \leq W_2(\mu,\pi)\sqrt{\langle f, \mathcal{L}\log f\rangle_\pi} - \frac{c}{2}W_2^2(\mu,\pi). \tag{56}$$

To emphasis the importance of this inequality let us first show that the result of Bakry and Emery (Theorem 5.7) easily follows follows from the HWI inequality. Using

$$W_2(\mu,\pi)\sqrt{\langle f, \mathcal{L}\log f\rangle_\pi} \leq \frac{c}{2}W_2^2(\mu,\pi) + \frac{1}{2c}\langle f, \mathcal{L}\log f\rangle_\pi,$$

in (56) we find that

$$\mathrm{Ent}_\pi(f) \leq \frac{c}{2}W_2^2(\mu,\pi) + \frac{1}{2c}\langle f, \mathcal{L}\log f\rangle_\pi - \frac{c}{2}W_2^2(\mu,\pi) = \frac{1}{2c}\langle f, \mathcal{L}\log f\rangle_\pi.$$

Then replacing $f = g^2$ gives the desired result.

□

Now we move to the proof of Theorem 12.1. To this end, we use the geometric picture that was developed in the previous section. We start by computing the second derivative of the entropy function. Let $\{\mu_t : t \geq 0\}$ be an arbitrary curve of probability measures with $\partial_t\mu_t = -\mathrm{div}(\mu_t\nabla\psi_t)$. Also let $f_t\mathrm{d}\pi = \mathrm{d}\mu_t$. Then we have

$$\begin{aligned}\partial_t \,\mathrm{Ent}_\pi(f_t) &= -\langle \mathrm{div}(\mu_t\nabla\psi_t), \log\mu_t - V\rangle \\ &= \langle \mu_t\nabla\psi_t, \nabla(\log\mu_t) + \nabla V\rangle \\ &= \langle \nabla\psi_t, \nabla\mu_t\rangle + \langle \mu_t\nabla\psi_t, \nabla V\rangle.\end{aligned} \tag{57}$$

Next we have

$$\begin{aligned}\partial_t^2 \,\mathrm{Ent}_\pi(f_t) &= \langle \nabla\partial_t\psi_t, \nabla\mu_t\rangle + \langle \nabla\psi_t, \nabla\partial_t\mu_t\rangle + \langle \partial_t(\mu_t\nabla\psi_t), \nabla V\rangle \\ &= -\langle \partial_t\psi_t, \Delta\mu_t\rangle - \langle \Delta\psi_t, \partial_t\mu_t\rangle + \langle \partial_t\mu_t\nabla\psi_t, \nabla V\rangle + \langle \mu_t\nabla\partial_t\psi_t, \nabla V\rangle \\ &= -\langle \partial_t\psi_t, \Delta\mu_t\rangle + \langle \Delta\psi_t, \mathrm{div}(\mu_t\nabla\psi_t)\rangle - \langle \mathrm{div}(\mu_t\nabla\psi_t)\nabla\psi_t, \nabla V\rangle - \langle \partial_t\psi_t, \mathrm{div}(\mu_t\nabla V)\rangle \\ &= -\langle \partial_t\psi_t, \Delta\mu_t + \mathrm{div}(\mu_t\nabla V)\rangle + \langle \Delta\psi_t, \mathrm{div}(\mu_t\nabla\psi_t)\rangle - \langle \mathrm{div}(\mu_t\nabla\psi_t), \nabla V\cdot\nabla\psi_t\rangle \\ &= -\langle \partial_t\psi_t, \Delta\mu_t + \mathrm{div}(\mu_t\nabla V)\rangle - \langle \mathcal{L}\psi_t, \mathrm{div}(\mu_t\nabla\psi_t)\rangle.\end{aligned}$$

Now let us further assume that $\{\mu_t : t \geq 0\}$ is a geodesic. Then by the geodesic equation (69) we find that

$$\begin{aligned}\partial_t^2 \,\mathrm{Ent}_\pi(f_t) &= \frac{1}{2}\langle |\nabla\psi_t|^2, \Delta\mu_t + \mathrm{div}(\mu_t\nabla V)\rangle - \langle \mathcal{L}\psi_t, \mathrm{div}(\mu_t\nabla\psi_t)\rangle \\ &= -\frac{1}{2}\langle |\nabla\psi_t|^2, \mathcal{L}^*\mu_t\rangle - \langle \mathcal{L}\psi_t, \mathrm{div}(\mu_t\nabla\psi_t)\rangle,\end{aligned} \tag{58}$$

where as before $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$ with respect to the inner product (50) given by

$$\mathcal{L}^*\mu = -\Delta\mu - \mathrm{div}(\mu\nabla V).$$

Let $B(\mu_t, \psi_t)$ be the right hand side of (58), i.e.,

$$B(\mu,\psi) = -\frac{1}{2}\langle |\nabla\psi|^2, \mathcal{L}^*\mu\rangle - \langle \mathcal{L}\psi, \mathrm{div}(\mu\nabla\psi)\rangle. \tag{59}$$

Lemma 12.2. *Let π be a Borel probability measure on $\mathbb{R}^k$ with $\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x$ where $\mathrm{Hess}(V) \geq cI$ for some constant c. Then the entropy function (with respect to π) is c-convex. In other words, for any geodesic $\{\mu_t : t \geq 0\}$ we have*

$$\partial_t^2 \operatorname{Ent}_\pi(f_t) \geq c \int_{\mathbb{R}^k} |\nabla \psi_t|^2 \mathrm{d}\mu_t,$$

where f_t, ψ_t are given by $f_t \pi = \mu$ and $\partial_t \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t)$.

Proof. By (58) and (59) it suffices to show that $B(\mu, \psi) \geq c|\nabla \psi|^2$. Observe that

$$B(\mu, \psi) = -\frac{1}{2}\langle \mathcal{L}|\nabla\psi|^2, \mu\rangle + \langle \nabla\psi \cdot \nabla\mathcal{L}\psi, \mu\rangle = \langle \mu, -\frac{1}{2}\mathcal{L}|\nabla\psi|^2 + \nabla\psi \cdot \nabla\mathcal{L}\psi\rangle.$$

We then compute

$$\begin{aligned}
\nabla\psi \cdot \nabla(\mathcal{L}\psi) &= \nabla\psi \cdot \nabla(\nabla V \cdot \nabla\psi - \Delta\psi) \\
&= \sum_{i,j} \partial_i\partial_j V \partial_i\psi\partial_j\psi + \partial_i V \partial_i\partial_j\psi\partial_j\psi - \partial_j\psi\partial_j\partial_i^2\psi, \\
&= \mathrm{Hess}(V)(\nabla\psi, \nabla\psi) + \mathrm{Hess}(\psi)(\nabla\psi, \nabla V) - \sum_{i,j} \partial_j\psi\partial_j\partial_i^2\psi,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{L}|\nabla\psi|^2 &= \nabla V \cdot \nabla|\nabla\psi|^2 - \Delta|\nabla\psi|^2 \\
&= 2\sum_{i,j} \partial_j V \partial_i\psi\partial_i\partial_j\psi - (\partial_i\partial_j\psi)^2 - \partial_i\psi\partial_i\partial_j^2\psi \\
&= 2\Big(\mathrm{Hess}(\psi)(\nabla\psi, \nabla V) - \sum_{i,j}(\partial_i\partial_j\psi)^2 + \partial_i\psi\partial_i\partial_j^2\psi\Big).
\end{aligned}$$

Therefore

$$-\frac{1}{2}\mathcal{L}|\nabla\psi|^2 + \nabla\psi \cdot \nabla\mathcal{L}\psi = \mathrm{Hess}(V)(\nabla\psi, \nabla\psi) + \sum_{i,j}(\partial_i\partial_j\psi)^2 \geq \mathrm{Hess}(V)(\nabla\psi, \nabla\psi).$$

Then by the assumption $\mathrm{Hess}(V) \geq cI$ we have

$$B(\mu, \psi) \geq \langle \mu, c|\nabla\psi|^2\rangle = c\int_{\mathbb{R}^k} |\nabla\psi|^2 \mathrm{d}\mu. \tag{60}$$

□

We need yet another lemma to prove the HWI inequality.

Lemma 12.3. *Let $\{\mu^s : 0 \leq s \leq 1\}$ be a geodesic with $f^s \mathrm{d}\pi = \mathrm{d}\mu^s$. Define*

$$f_t^s = e^{-st\mathcal{L}} f^s,$$

and $\mathrm{d}\mu_t^s = f_t^s \mathrm{d}\pi$ (so that $\mu_t^s = e^{-st\mathcal{L}^*}\mu^s$). Define ψ_t^s by

$$\partial_s \mu_t^s = -\mathrm{div}(\mu_t^s \nabla \psi_t^s).$$

Then we have

$$\frac{1}{2}\partial_t \int_{\mathbb{R}^k} |\nabla\psi_t^s|^2 \mathrm{d}\mu_t^s + \partial_s \operatorname{Ent}_\pi(f_t^s) = -sB(\mu_t^s, \psi_t^s).$$

The proof of this lemma is left for Appendix E. Now we are ready to prove the main result of this section that is a stronger version of the HWI inequality.

Theorem 12.4. *Let π be a Borel probability measure on $\mathbb{R}^k$ with $\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x$ where $\mathrm{Hess}(V) \geq cI$ for some constant c. Let $\mu_t = e^{-t\mathcal{L}^*}\mu_0$ and define f_t by $f_t\pi = \mu_t$ so that $f_t = e^{-t\mathcal{L}}f_0$. Also let τ be another probability measure with $\tau = g\pi$. Then we have*

$$\frac{1}{2}\partial_t W_2^2(\mu_t, \tau)\Big|_{t=0} + \frac{c}{2}W_2^2(\mu_0, \tau) \leq \mathrm{Ent}_\pi(g) - \mathrm{Ent}_\pi(f_0).$$

Before proving this theorem let us first show how the HWI inequality is derived from it.

Proof of Theorem 12.1. In the statement of Theorem 12.4 let $\tau = \pi$. Then we have

$$\frac{1}{2}\partial_t W_2^2(\mu_t, \pi)\Big|_{t=0} + \frac{c}{2}W_2^2(\mu_0, \pi) \leq -\mathrm{Ent}_\pi(f_0). \tag{61}$$

On the other hand, by the triangle inequality we have $W_2(\mu_0, \pi) \leq W_2(\mu_t, \mu_0) + W_2(\mu_t, \pi)$. Raising both sides to the power of two, we find that

$$W_2^2(\mu_t, \pi) - W_2^2(\mu_0, \pi) \leq -W_2^2(\mu_t, \mu_0) - 2W_2(\mu_t, \mu_0)W_2(\mu_t, \pi).$$

Therefore,

$$\begin{aligned}
\frac{1}{2}\partial_t W_2^2(\mu_t, \pi)\Big|_{t=0} &\geq \lim_{t\to 0^+}\frac{1}{2t}\Big(- W_2^2(\mu_t, \mu_0) - 2W_2(\mu_t, \mu_0)W_2(\mu_t, \pi)\Big) \\
&= -\lim_{t\to 0^+}\frac{1}{2t}W_2^2(\mu_t, \mu_0) - W_2(\mu_0, \pi)\lim_{t\to 0^+}\frac{1}{t}W_2(\mu_t, \mu_0) \\
&\overset{(a)}{=} -\lim_{t\to 0^+} W_2(\mu_t, \mu_0)\partial_t W_2(\mu_t, \mu_0) - W_2(\mu_0, \pi)\partial_t W_2(\mu_t, \mu_0)\Big|_{t=0} \\
&\overset{(b)}{\geq} -\lim_{t\to 0^+} W_2(\mu_t, \mu_0)\sqrt{\langle f_t, \mathcal{L}\log f_t\rangle_\pi} - W_2(\mu_0, \pi)\sqrt{\langle f_0, \mathcal{L}\log f_0\rangle_\pi} \\
&= -W_2(\mu_0, \pi)\sqrt{\langle f_0, \mathcal{L}\log f_0\rangle_{\mu_0}}.
\end{aligned}$$

Here for (a) we use L'Hôpital's rule and for (b) we use Lemma 11.1 twice. Using this inequality in (61), the desired result follows. $\square$

We now prove Theorem 12.4.

Proof of Theorem 12.4. Let $\mu^0 = \tau$ and $\mu^1 = \mu_0$ and let $\{\mu^s : 0 \leq s \leq 1\}$ be a geodesic between μ^0, μ^1. Also define

$$\mu_t^s = e^{-st\mathcal{L}^*}\mu^s,$$

and f_t^s as before. Observe that $\mu_t^1 = \mu_t$ and $f_t^1 = f_t$. By Lemma 12.2 (equation (60)) and Lemma 12.3 we have

$$\frac{1}{2}\partial_t\langle\nabla\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s} + \partial_s \mathrm{Ent}_\pi(f_t^s) \leq -cs\langle\nabla\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s}.$$

Multiplying both sides by e^{2cst} we obtain

$$\frac{1}{2}\partial_t\Big(e^{2cst}\langle\nabla\psi_t^s,\nabla\psi_t^s\rangle_{\mu_t^s}\Big)+e^{2cst}\partial_s\operatorname{Ent}_\pi(f_t^s)\le 0,$$

or equivalently

$$\frac{1}{2}\partial_t\Big(e^{2cst}\langle\nabla\psi_t^s,\nabla\psi_t^s\rangle_{\mu_t^s}\Big)+\partial_s\Big(e^{2cst}\operatorname{Ent}_\pi(f_t^s)\Big)\le 2cte^{2cst}\operatorname{Ent}_\pi(f_t^s).$$

Integrating t over $[0,\epsilon]$ and s over $[0,1]$ we find that

$$\begin{aligned}\frac{1}{2}\int_0^1\Big(e^{2cs\epsilon}\langle\nabla\psi_\epsilon^s,\nabla\psi_\epsilon^s\rangle_{\mu_\epsilon^s}-\langle\nabla\psi_0^s,\nabla\psi_0^s\rangle_{\mu_0^s}\Big)\mathrm{d}s+\int_0^\epsilon\Big(e^{2ct}\operatorname{Ent}_\pi(f_t^1)-\operatorname{Ent}_\pi(f_t^0)\Big)\mathrm{d}t&\\ \le 2c\int_0^1\int_0^\epsilon te^{2cst}\operatorname{Ent}_\pi(f_t^s)\mathrm{d}s\mathrm{d}t&\end{aligned}$$

On the other hand, by a re-parametrization argument (see below) we obtain

$$\frac{1}{2}\int_0^1\Big(e^{2cs\epsilon}\langle\nabla\psi_\epsilon^s,\nabla\psi_\epsilon^s\rangle_{\mu_\epsilon^s}-\langle\nabla\psi_0^s,\nabla\psi_0^s\rangle_{\mu_0^s}\Big)\mathrm{d}s\ge\frac{c\epsilon}{1-e^{-2c\epsilon}}W_2^2(\mu_\epsilon,\tau)-\frac{1}{2}W_2^2(\mu_0,\tau).\quad(62)$$

On the other hand, by the data processing inequality we have

$$\operatorname{Ent}_\pi(f_t^0)\le\operatorname{Ent}_\pi(f^0)=\operatorname{Ent}_\pi(g).$$

Putting these together we arrive at

$$\begin{aligned}\frac{c\epsilon}{1-e^{-2c\epsilon}}W_2^2(\mu_\epsilon,\tau)-\frac{1}{2}W_2^2(\mu_0,\tau)+\int_0^\epsilon e^{2ct}\operatorname{Ent}_\pi(f_t^1)\mathrm{d}t-\epsilon\operatorname{Ent}_\pi(g)&\\ \le 2c\int_0^1\int_0^\epsilon te^{2cst}\operatorname{Ent}_\pi(f_t^s)\mathrm{d}s\mathrm{d}t.&\end{aligned}$$

Once again by the data processing inequality we have $\operatorname{Ent}_\pi(f_t^s)\le\operatorname{Ent}_\pi(f^s)$. Therefore,

$$\begin{aligned}\int_0^\epsilon te^{2cst}\operatorname{Ent}_\pi(f_t^s)\mathrm{d}t&\le\int_0^\epsilon te^{2cst}\operatorname{Ent}_\pi(f^s)\mathrm{d}s\\ &\le e^{2cs}\operatorname{Ent}_\pi(f^s)\int_0^\epsilon t\mathrm{d}t\\ &=\frac{\epsilon^2}{2}e^{cs}\operatorname{Ent}_\pi(f^s),\end{aligned}$$

where in the second line we use $0\le t\le\epsilon\le 1$ so that $e^{2cst}\le e^{2cs}$. Therefore,

$$\frac{c\epsilon}{1-e^{-2c\epsilon}}W_2^2(\mu_\epsilon,\tau)-\frac{1}{2}W_2^2(\mu_0,\tau)+\int_0^\epsilon e^{2ct}\operatorname{Ent}_\pi(f_t^1)\mathrm{d}t-\epsilon\operatorname{Ent}_\pi(g)\le c\epsilon^2\int_0^1 e^{cs}\operatorname{Ent}_\pi(f^s)\mathrm{d}s.$$

Next, dividing both sides by ϵ and taking the limit $\epsilon\to 0$ we obtain

$$\partial_\epsilon\Big(\frac{c\epsilon}{1-e^{-2c\epsilon}}W_2^2(\mu_\epsilon,\tau)\Big)\Big|_{\epsilon=0}+\operatorname{Ent}_\pi(f_0)-\operatorname{Ent}_\pi(g)\le 0,$$

where we use $\lim_{\epsilon\to 0}\frac{c\epsilon}{1-e^{-2c\epsilon}} = 1/2$. Also, using $\partial_\epsilon\big(\frac{c\epsilon}{1-e^{-2c\epsilon}}\big)\big|_{\epsilon=0} = c/2$ we find that

$$\frac{1}{2}\partial_\epsilon W_2^2(\mu_\epsilon,\tau) + \Big|_{\epsilon=0} + \frac{c}{2}W_2^2(\mu_0,\tau) + \mathrm{Ent}_\pi(f_0) - \mathrm{Ent}_\pi(g) \le 0,$$

that is equivalent to the desired result.

It is remained to prove (62). Let $\theta : [0,1]\to[0,1]$ be an arbitrary increasing function with $\theta(0)=0$ and $\theta(1)=1$. Consider the curve $\{\mu_\epsilon^{\theta(r)} : 0\le r\le 1\}$. We have

$$\mu_\epsilon^{\theta(0)} = \mu_\epsilon^0 = \mu^0 = \tau, \qquad \mu_\epsilon^{\theta(1)} = \mu_\epsilon^1 = \mu_\epsilon.$$

Thus we have a curve between τ and μ_ϵ. We also have

$$\partial_r \mu_\epsilon^{\theta(r)} = -\theta'(r)\mathrm{div}(\mu_\epsilon^{\theta(r)}\nabla\psi_\epsilon^{\theta(r)}).$$

Therefore, by the the Benamou-Brenier formula we have

$$W_2^2(\mu_\epsilon,\tau) \le \int_0^1 \theta'(r)^2\langle\nabla\psi_\epsilon^{\theta(r)},\nabla\psi_\epsilon^{\theta(r)}\rangle_{\mu_\epsilon^{\theta(r)}}\mathrm{d}r.$$

Letting $r=\theta^{-1}(s)$ we have

$$W_2^2(\mu_\epsilon,\tau) \le \int_0^1 \theta'(\theta^{-1}(s))\langle\nabla\psi_\epsilon^s,\nabla\psi_\epsilon^s\rangle_{\mu_\epsilon^s}\mathrm{d}s.$$

Now pick

$$\theta^{-1}(s) = \frac{1}{m_\epsilon}\int_0^s e^{-2c\epsilon v}\mathrm{d}v,$$

where

$$m_\epsilon = \int_0^1 e^{-2c\epsilon v}\mathrm{d}v = \frac{1-e^{-2c\epsilon}}{2c\epsilon},$$

is a normalization constant so that $\theta(1)=1$. Then we have $\theta'(\theta^{-1}(s)) = m_\epsilon e^{2c\epsilon s}$ and

$$W_2^2(\mu_\epsilon,\tau) \le m_\epsilon\int_0^1 e^{2c\epsilon s}\langle\nabla\psi_\epsilon^s,\nabla\psi_\epsilon^s\rangle_{\mu_\epsilon^s}\mathrm{d}s.$$

Therefore, using the fact that $\{\mu^s : 0\le s\le 1\}$ is a geodesic we have

$$\frac{1}{2}\int_0^1\Big(e^{2cs\epsilon}\langle\nabla\psi_\epsilon^s,\nabla\psi_\epsilon^s\rangle_{\mu_\epsilon^s} - \langle\nabla\psi_0^s,\nabla\psi_0^s\rangle_{\mu_0^s}\Big)\mathrm{d}s \ge \frac{1}{2m_\epsilon}W_2^2(\mu_\epsilon,\tau) - \frac{1}{2}W_2^2(\mu,\tau).$$

We are done.

□

We finish this section by two interesting consequences of Theorem 12.4. We start by the simpler one which is the exponential convergence rate in the 2-Wasserstein distance.

Corollary 12.5. *Let π be a Borel probability measure on $\mathbb{R}^k$ with $\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x$ where $\mathrm{Hess}(V)\ge cI$ for some constant c. Also let $\mu_t = e^{-t\mathcal{L}^*}\mu_0$ with $\mathrm{d}\mu_t = f_t\mathrm{d}\pi$. Then we have*

$$W_2(\mu_t,\pi) \le e^{-ct}W_2(\mu_0,\pi).$$

Proof. By Theorem 12.4 for the choice of $\tau = \pi$ we have

$$\frac{1}{2}\partial_t W_2^2(\mu_t, \pi) + \frac{c}{2} W_2^2(\mu_t, \pi) \leq -\operatorname{Ent}_\pi(f_t).$$

Now as shown in the beginning of this section the assumption $\operatorname{Hess}(V) \geq cI$ gives a log-Sobolev inequality (the Bakry-Emery result) which itself gives the following 2-Talagrand inequality

$$\frac{c}{2} W_2^2(\mu_t, \pi) \leq \operatorname{Ent}_\pi(\mu_t).$$

Using this in the previous inequality we obtain

$$\frac{1}{2}\partial_t W_2^2(\mu_t, \pi) \leq -c W_2^2(\mu_t, \pi),$$

and then

$$W_2^2(\mu_t, \pi) \leq e^{-2ct} W_2^2(\mu_t, \pi),$$

which is equivalent to what we want.

□

Corollary 12.6. *Let π be a Borel probability measure on $\mathbb{R}^k$ with $\mathrm{d}\pi = e^{-V(\mathbf{x})}\mathrm{d}x$ where $\operatorname{Hess}(V) \geq cI$ for some constant c. Then the entropy function is geodesically c-convex. Namely, for any geodesic $\{\mu^s : 0 \leq s \leq 1\}$ with $\mu^s = f^s \pi$ we have*

$$(1-s)\operatorname{Ent}_\pi(f^0) + s\operatorname{Ent}_\pi(f^1) - \operatorname{Ent}_\pi(f^s) \geq \frac{c}{2} s(1-s) W_2^2(\mu^0, \mu^1), \qquad \forall s.$$

Proof. Let $\mu_t^s = e^{-t\mathcal{L}^*}\mu^s$ and $f_t^s = e^{-t\mathcal{L}} f^s$. For any $s \in [0,1]$ and $j \in \{0,1\}$ define

$$h(t) = \frac{e^{ct}}{2} W_2^2(\mu_t^s, \mu^j) + \int_0^t e^{cr}\big(\operatorname{Ent}_\pi(f_r^s) - \operatorname{Ent}_\pi(f^j)\big)\mathrm{d}r.$$

Then computing the derivative of $h(t)$ and using Theorem 12.4 we find that $h(t)$ is non-increasing. Therefore, $h(t) \leq h(0)$, i.e.,

$$\begin{aligned}
\frac{e^{ct}}{2} W_2^2(\mu_t^s, \mu^j) - \frac{1}{2} W_2^2(\mu^s, \mu^j) &\leq -\int_0^t e^{cr}\Big(\operatorname{Ent}_\pi(f_r^s) - \operatorname{Ent}_\pi(f^j)\Big)\mathrm{d}r \\
&\leq -\int_0^t e^{cr}\Big(\operatorname{Ent}_\pi(f_t^s) - \operatorname{Ent}_\pi(f^j)\Big)\mathrm{d}r \\
&= \frac{e^{ct}-1}{c}\Big(\operatorname{Ent}_\pi(f^j) - \operatorname{Ent}_\pi(f_t^s)\Big),
\end{aligned}$$

where for the second inequality we use the data processing inequality $\operatorname{Ent}_\pi(f_r^s) \geq \operatorname{Ent}_\pi(f_t^s)$ as $r \leq t$. Next taking the average of these two inequalities for the choices of $j = 0, 1$ we obtain

$$\begin{aligned}
\frac{e^{ct}}{2}\Big((1-s)W_2^2(\mu_t^s, \mu^0) + sW_2^2(\mu_t^s, \mu^1)\Big) &- \frac{1}{2}\Big((1-s)W_2^2(\mu^s, \mu^0) + sW_2^2(\mu^s, \mu^1)\Big) \\
&\leq \frac{e^{ct}-1}{c}\Big((1-s)\operatorname{Ent}_\pi(f^0) + s\operatorname{Ent}_\pi(f^1) - \operatorname{Ent}_\pi(f_t^s)\Big),
\end{aligned}$$

Now using $(1-s)a^2 + sb^2 \geq s(1-s)(a+b)^2$ that holds for all $0 \leq s \leq 1$ we have

$$\begin{aligned}(1-s)W_2^2(\mu_t^s, \mu^0) + sW_2^2(\mu_t^s, \mu^1) &\geq s(1-s)\Big(W_2(\mu_t^s, \mu^0) + W_2(\mu_t^s, \mu^1)\Big)^2 \\ &\geq s(1-s)W_2^2(\mu^0, \mu^1).\end{aligned}$$

Moreover, since $\{\mu^s : 0 \leq s \leq 1\}$ is a geodesic we have

$$\begin{aligned}(1-s)W_2^2(\mu^s, \mu^0) + sW_2^2(\mu^s, \mu^1) &= (1-s)s^2W_s^2(\mu^0, \mu^1) + s(1-s)^2W_2^2(\mu^0, \mu^1) \\ &= s(1-s)W_s^2(\mu^0, \mu^1).\end{aligned}$$

We conclude that

$$\frac{e^{ct}-1}{2}s(1-s)W_2^2(\mu^0, \mu^1) \leq \frac{e^{ct}-1}{c}\Big((1-s)\operatorname{Ent}_\pi(f^0) + s\operatorname{Ent}_\pi(f^1) - \operatorname{Ent}_\pi(f_t^s)\Big),$$

or equivalently

$$\frac{c}{2}s(1-s)W_2^2(\mu^0, \mu^1) \leq \Big((1-s)\operatorname{Ent}_\pi(f^0) + s\operatorname{Ent}_\pi(f^1) - \operatorname{Ent}_\pi(f_t^s)\Big).$$

Letting $t = 0$ we obtain the desired result.

$\square$

Appendix

A A proof of Theorem 5.7

Let $f_t = T_t f$ with $\{T_t : t \geq 0\}$ being the semigroup generated by $\mathcal{L} = \nabla V \cdot \nabla - \Delta$. Also let $g_t = \sqrt{f_t}$. Then we have $\partial_t f_t = -\mathcal{L} f_t$ where $\partial_t = \partial/\partial t$. A simple computation verifies that

$$\partial_t \operatorname{Ent}_\pi(f_t) = -4\langle g_t, \mathcal{L} g_t\rangle_\pi.$$

Also it can be shown that

$$\partial_t g_t = -\mathcal{L} g_t + \frac{|\nabla g_t|^2}{g_t}.$$

Next we compute

$$\partial_t \langle g_t, \mathcal{L} g_t\rangle_\pi = -2\langle \mathcal{L} g_t, \mathcal{L} g_t\rangle_\pi + 2\Big\langle \frac{|\nabla g_t|^2}{g_t}, \mathcal{L} g_t\Big\rangle_\pi. \tag{63}$$

The second term equals

$$
\begin{aligned}
\Big\langle \frac{|\nabla g_t|^2}{g_t}, \mathcal{L} g_t \Big\rangle_\pi &= \Big\langle \nabla \frac{|\nabla g_t|^2}{g_t}, \nabla g_t \Big\rangle_\pi \\
&= \Big\langle \nabla g_t, \nabla(\frac{1}{g_t})|\nabla g_t|^2 + \frac{1}{g_t}\nabla|\nabla g_t|^2 \Big\rangle_\pi \\
&= \Big\langle \nabla g_t, -\frac{1}{g_t^2}\nabla(g_t)|\nabla g_t|^2 + \frac{1}{g_t}\nabla|\nabla g_t|^2 \Big\rangle_\pi \\
&= \int_{\mathbb{R}^k} \Big[-\frac{(\partial_i g_t)^2(\partial_j g_t^2)}{g_t^2} + 2\frac{\partial_i g_t \partial_j g_t \partial_i \partial_j g_t}{g_t} \Big] \mathrm{d}\pi,
\end{aligned}
$$

where we suppress a summation over i, j. Also for the first term in (63) we have

$$
\begin{aligned}
\langle \mathcal{L} g_t, \mathcal{L} g_t \rangle_\pi &= \langle \nabla \mathcal{L} g_t, \nabla g_t \rangle_\pi \\
&= \langle \nabla(\nabla V \cdot \nabla g_t - \Delta g_t), \nabla g_t \rangle_\pi \\
&= \int_{\mathbb{R}^k} \Big[\partial_i g_t \partial_i (\partial_j V \partial_j g_t) - \partial_i g_t \partial_i (\partial_j^2 g_t) \Big] \mathrm{d}\pi \\
&= \int_{\mathbb{R}^k} \Big[\partial_i g_t (\partial_i \partial_j V \partial_j g_t + \partial_j V \partial_i \partial_j g_t) - \partial_i g_t \partial_i (\partial_j^2 g_t) \Big] \mathrm{d}\pi.
\end{aligned}
$$

On the other hand we have

$$
\begin{aligned}
\int_{\mathbb{R}^k} \partial_i g_t \partial_i (\partial_j^2 g_t) \mathrm{d}\pi &= \int_{\mathbb{R}^k} \partial_i g_t \partial_i (\partial_j^2 g_t) e^{-V} \mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^k} \partial_j \big(e^{-V} \partial_i g_t \big) \partial_i \partial_j g_t \mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^k} \Big[-\partial_j V e^{-V} \partial_i g_t \partial_i \partial_j g_t + e^{-V} (\partial_i \partial_j g_t)^2 \Big] \mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^k} \Big[-\partial_j V \partial_i g_t \partial_i \partial_j g_t + (\partial_i \partial_j g_t)^2 \Big] \mathrm{d}\pi.
\end{aligned}
$$

Therefore,

$$
\langle \mathcal{L} g_t, \mathcal{L} g_t \rangle_\pi = \int_{\mathbb{R}^k} \Big[\partial_i g_t \partial_i \partial_j V \partial_j g_t + (\partial_i \partial_j g_t)^2 \Big] \mathrm{d}\pi.
$$

Putting these in (63) we find that

$$
\begin{aligned}
\partial_t \langle g_t, \mathcal{L} g_t \rangle_\pi &= -2 \int_{\mathbb{R}^k} \Big[\partial_i g_t \partial_i \partial_j V \partial_j g_t + (\partial_i \partial_j g_t)^2 + \frac{(\partial_i g_t)^2 (\partial_j g_t)^2}{g_t^2} - 2\frac{\partial_i g_t \partial_j g_t \partial_i \partial_j g_t}{g_t} \Big] \mathrm{d}\pi \\
&= -2 \int_{\mathbb{R}^k} \Big[\mathrm{Hess}(V)(\nabla g_t, \nabla g_t) + \Big(\partial_i \partial_j g_t - \frac{\partial_i g_t \partial_j g_t}{g_t} \Big)^2 \Big] \mathrm{d}\pi \\
&\leq -2 \int_{\mathbb{R}^k} \mathrm{Hess}(V)(\nabla g_t, \nabla g_t) \mathrm{d}\pi \\
&\leq -2c \langle \nabla g_t, \nabla g_t \rangle_\pi \\
&= -2c \langle g_t, \mathcal{L} g_t \rangle_\pi,
\end{aligned}
$$

where in the last inequality we use the assumption $\mathrm{Hess}(V) \geq cI$. We conclude that

$$\langle g_t, \mathcal{L} g_t\rangle_\pi \leq e^{-2ct}\langle g_0, \mathcal{L} g_0\rangle_\pi.$$

Therefore we have

$$\begin{aligned}
\mathrm{Ent}_\pi(g_0^2) &= \mathrm{Ent}_\pi(f_0) \\
&= -\int_0^\infty \partial_t \mathrm{Ent}_\pi(f_t)\mathrm{d}t \\
&= 4\int_0^\infty \langle g_t, \mathcal{L} g_t\rangle_\pi \mathrm{d}t \\
&\leq 4\int_0^\infty e^{-2ct}\langle g_0, \mathcal{L} g_0\rangle_\pi \mathrm{d}t \\
&= \frac{2}{c}\langle g_0, \mathcal{L} g_0\rangle_\pi.
\end{aligned}$$

We are done.

B Proof of Lemma 8.2

Define

$$g(x) = \ln\big((1-p)e^{-px} + pe^{(1-p)x}\big).$$

It is easy to verify that $g(0) = g'(0) = 0$ and

$$g''(x) = \frac{p(1-p)e^x}{\big(pe^x + (1-p)\big)^2} \leq \frac{1}{4}, \qquad \forall x.$$

Then by Taylor's theorem for some ξ between 0 and x we have

$$g(x) = g(0) + g'(0)x + g''(\xi)\frac{x^2}{2} \leq \frac{x^2}{8}.$$

We are done.

C Hopf-Lax formula

We will show that given $f : \mathbb{R}^k \to \mathbb{R}$ the function $(t, \mathbf{x}) \mapsto Q_t f(\mathbf{x})$ defined by

$$Q_t f(\mathbf{x}) = \inf_{\mathbf{y}\in\mathbb{R}^k} f(\mathbf{y}) + \frac{1}{t}|\mathbf{x}-\mathbf{y}|^2,\ t > 0, \qquad Q_0 f = f, \tag{64}$$

gives a solution of the Hamilton-Jacobi equation:

$$\begin{cases} \frac{\partial}{\partial t}u + \frac{1}{4}|\nabla_{\mathbf{x}} u|^2 = 0, & t \geq 0, \mathbf{x} \in \mathbb{R}^k \\ u(0,\mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^k. \end{cases} \tag{65}$$

The facts that $\lim_{t\to 0^+} Q_t f(\mathbf{x}) = f(\mathbf{x})$ and that $Q_t f(\mathbf{x})$ is almost everywhere differentiable (i.e., it is a Lipschitz function) is left to the reader. So we verify (65). Before that it is instructive to note that $\{Q_t : t \geq 0\}$ forms a semigroup. To show this we compute

$$\begin{aligned} Q_s Q_t f(\mathbf{x}) &= \inf_{\mathbf{y}} Q_t f(\mathbf{y}) + \frac{1}{s}|\mathbf{y}-\mathbf{x}|^2 \\ &= \inf_{\mathbf{y}} \inf_{\mathbf{z}} f(\mathbf{z}) + \frac{1}{t}|\mathbf{y}-\mathbf{z}|^2 + \frac{1}{s}|\mathbf{y}-\mathbf{x}|^2 \\ &= \inf_{\mathbf{z}} f(\mathbf{z}) + \inf_{\mathbf{y}} \frac{1}{t}|\mathbf{y}-\mathbf{z}|^2 + \frac{1}{s}|\mathbf{y}-\mathbf{x}|^2 \\ &= \inf_{\mathbf{z}} f(\mathbf{z}) + \frac{1}{s+t}|\mathbf{x}-\mathbf{z}|^2 \\ &= Q_{s+t} f(\mathbf{x}), \end{aligned}$$

where we used

$$\inf_{\mathbf{y}} \frac{1}{t}|\mathbf{y}-\mathbf{z}|^2 + \frac{1}{s}|\mathbf{y}-\mathbf{x}|^2 = \frac{1}{s+t}|\mathbf{x}-\mathbf{z}|^2,$$

which is an easy exercise to prove.

We now turn to the proof of (65). For every $\mathbf{r} \in \mathbb{R}^k$ and $\epsilon > 0$ we have

$$\begin{aligned} Q_{t+\epsilon} f(\mathbf{x}+\epsilon\mathbf{r}) &= Q_\epsilon Q_t f(\mathbf{x}+\epsilon\mathbf{r}) \\ &= \inf_{\mathbf{y}} Q_t f(\mathbf{y}) + \frac{1}{\epsilon}|\mathbf{x}-\mathbf{y}+\epsilon\mathbf{r}|^2 \\ &\leq Q_t f(\mathbf{x}) + \epsilon|\mathbf{r}|^2. \end{aligned}$$

Dividing both sides by ϵ and taking the limit $\epsilon \to 0^+$ gives

$$\frac{\partial}{\partial t} Q_t f(\mathbf{x}) + \mathbf{r} \cdot \nabla Q_t f(\mathbf{x}) \leq |\mathbf{r}|^2.$$

Now optimizing over the choice of $\mathbf{r} \in \mathbb{R}^k$ we arrive at

$$\frac{\partial}{\partial t} Q_t f(\mathbf{x}) + \frac{1}{4}\left|\nabla Q_t f(\mathbf{x})\right|^2 \leq 0.$$

To prove inequality in the other direction we just need to find some $\mathbf{r}_0 \in \mathbb{R}^k$ such that

$$\frac{\partial}{\partial t} Q_t f(\mathbf{x}) + \mathbf{r}_0 \cdot \nabla Q_t f(\mathbf{x}) \geq |\mathbf{r}_0|^2. \tag{66}$$

In this case we would have

$$\frac{\partial}{\partial t} Q_t f(\mathbf{x}) \geq |\mathbf{r}_0|^2 - \mathbf{r}_0 \cdot \nabla Q_t f(\mathbf{x}) \geq \inf_{\mathbf{r}} |\mathbf{r}|^2 - \mathbf{r} \cdot \nabla Q_t f(\mathbf{x}) = -\frac{1}{4}\left|\nabla Q_t f(\mathbf{x})\right|^2.$$

To find such $\mathbf{r}_0$ let us take $\mathbf{z}$ to be such that

$$Q_t f(\mathbf{x}) = f(\mathbf{z}) + \frac{1}{t}|\mathbf{x}-\mathbf{z}|^2.$$

For arbitrary $\epsilon > 0$ define

$$\mathbf{r}_0 = \frac{\mathbf{x} - \mathbf{z}}{t}, \qquad \mathbf{y} = \mathbf{x} - \epsilon \mathbf{r}_0.$$

We now compute

$$\begin{aligned} Q_t f(\mathbf{x}) - Q_{t-\epsilon} f(\mathbf{x} - \epsilon \mathbf{r}_0) &= Q_t f(\mathbf{x}) - Q_{t-\epsilon} f(\mathbf{y}) \\ &\geq f(\mathbf{z}) + \frac{1}{t}|\mathbf{x} - \mathbf{z}|^2 - \big(f(\mathbf{z}) + \frac{1}{t-\epsilon}|\mathbf{y} - \mathbf{z}|^2\big) \\ &= \epsilon |\mathbf{r}_0|^2. \end{aligned}$$

Dividing both sides by ϵ and taking the limit $\epsilon \to 0^+$ we obtain the desired inequality (66).

D Geodesic equation of the 2-Wasserstein distance

Let μ_0, μ_1 be arbitrary probability measures. Then by the Benamou-Brenier formula we have

$$W_2^2(\mu_0, \mu_1) = \inf_{\{\mu_t : 0 \leq t \leq 1\}} \left\{ \int_0^1 \langle \nabla \psi_t, \nabla \psi_t \rangle_{\mu_t} \mathrm{d}t \,:\, \partial_t \mu_t = -\mathrm{div}(\mu_t \nabla \psi_t) \right\}.$$

Assume that $\{\mu_t : 0 \leq t \leq 1\}$ is a geodesic. Then for any set of functions $\{g_t : 0 \leq t \leq 1\}$ such that $\int_{\mathbb{R}^k} g_t \mathrm{d}\mathbf{x} = 0$ for all t and $g_0 = g_1 = 0$, the curve $\mu_t^\epsilon = \mu_t + \epsilon g_t$ is also a valid curve between μ_0, μ_1 in the space of probability measures assuming that $|\epsilon|$ is sufficiently small. Therefore, letting

$$\partial_t \mu_t^\epsilon = -\mathrm{div}(\mu_t^\epsilon \nabla \psi_t^\epsilon), \tag{67}$$

we must have

$$\partial_\epsilon \Big(\int_0^1 \langle \nabla \psi_t^\epsilon, \nabla \psi_t^\epsilon \rangle_{\mu_t^\epsilon} \mathrm{d}t \Big)\Big|_{\epsilon=0} = 0$$

Taking the derivative of both sides of (67) at $\epsilon = 0$ we find that

$$\partial_t g_t = -\mathrm{div}(g_t \nabla \psi_t + \mu_t V_t), \qquad Y_t = \partial_\epsilon (\nabla \psi_t^\epsilon)\Big|_{\epsilon=0}. \tag{68}$$

Therefore, we have

$$\begin{aligned}
0 &= \partial_\epsilon\Big(\int_0^1 \langle \nabla\psi_t^\epsilon, \nabla\psi_t^\epsilon\rangle_{\mu_t^\epsilon}\mathrm{d}t\Big)\Big|_{\epsilon=0} \\
&= 2\int_0^1 \langle Y_t, \nabla\psi_t\rangle_{\mu_t}\mathrm{d}t + \int_0^1 \langle \nabla\psi_t, \nabla\psi_t\rangle_{g_t}\mathrm{d}t \\
&= 2\int_0^1 \langle \mu_t Y_t, \nabla\psi_t\rangle\mathrm{d}t + \int_0^1 \langle g_t\nabla\psi_t, \nabla\psi_t\rangle\mathrm{d}t \\
&= -2\int_0^1 \langle \mathrm{div}(\mu_t Y_t), \psi_t\rangle\mathrm{d}t + \int_0^1 \langle g_t\nabla\psi_t, \nabla\psi_t\rangle\mathrm{d}t \\
&\overset{(a)}{=} 2\int_0^1 \langle \partial_t g_t + \mathrm{div}(g_t\nabla\psi_t), \psi_t\rangle\mathrm{d}t - \int_0^1 \langle \mathrm{div}(g_t\nabla\psi_t), \psi_t\rangle\mathrm{d}t \\
&= 2\int_0^1 \langle \partial_t g_t, \psi_t\rangle\mathrm{d}t + \int_0^1 \langle \mathrm{div}(g_t\nabla\psi_t), \psi_t\rangle\mathrm{d}t \\
&= -2\int_0^1 \langle g_t, \partial_t\psi_t\rangle\mathrm{d}t - \int_0^1 \langle g_t\nabla\psi_t, \nabla\psi_t\rangle\mathrm{d}t,
\end{aligned}$$

where in (a) we use (68). Since the above equation must hold for any choice functions $\{g_t : 0 \le t \le 1\}$ with the above constraint, we must have

$$\partial\psi_t + \frac{1}{2}|\nabla\psi_t|^2 = 0, \qquad \forall t. \tag{69}$$

E Proof of Lemma 12.3

We compute

$$\partial_t\langle \nabla\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s} = 2\langle \nabla\partial_t\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s} + \langle \nabla\psi_t^s, \nabla\psi_t^s\rangle_{\partial_t\mu_t^s}.$$

For the second term we have

$$\langle \nabla\psi_t^s, \nabla\psi_t^s\rangle_{\partial_t\mu_t^s} = \langle \partial_t\mu_t^s\nabla\psi_t^s, \nabla\psi_t^s\rangle = -\langle \mathcal{L}^*\mu_t^s\nabla\psi_t^s, \nabla\psi_t^s\rangle.$$

For the first term we have

$$\begin{aligned}
\langle \nabla\partial_t\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s} &= -\langle \partial_t\psi_t^s, \mathrm{div}(\mu_t^s\nabla\psi_t^s)\rangle \\
&= \langle \partial_t\psi_t^s, \partial_s\mu_t^s\rangle \\
&= \partial_t\langle \psi_t^s, \partial_s\mu_t^s\rangle - \langle \psi_t^s, \partial_s\partial_t\mu_t^s\rangle \\
&= -\partial_t\langle \psi_t^s, \mathrm{div}(\mu_t^s\nabla\psi_t^s)\rangle + \langle \psi_t^s, \partial_s(s\mathcal{L}^*\mu_t^s)\rangle \\
&= \partial_t\langle \nabla\psi_t^s, \mu_t^s\nabla\psi_t^s\rangle + \langle \psi_t^s, \partial_s(s\mathcal{L}^*\mu_t^s)\rangle.
\end{aligned}$$

Putting these together we arrive at

$$\partial_t\langle \nabla\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s} = 2\partial_t\langle \nabla\psi_t^s, \nabla\psi_t^s\rangle_{\mu_t^s} + 2\langle \psi_t^s, \partial_s(s\mathcal{L}^*\mu_t^s)\rangle - s\langle \mathcal{L}^*\mu_t^s\nabla\psi_t^s, \nabla\psi_t^s\rangle.$$

Therefore,

$$\begin{aligned}\frac{1}{2}\partial_t\langle\nabla\psi_t^s,\nabla\psi_t^s\rangle_{\mu_t^s} &= \frac{s}{2}\langle\mathcal{L}^*\mu_t^s\nabla\psi_t^s,\nabla\psi_t^s\rangle - \langle\psi_t^s,\partial_s(s\mathcal{L}^*\mu_t^s)\rangle \\ &= \frac{s}{2}\langle\mathcal{L}^*\mu_t^s\nabla\psi_t^s,\nabla\psi_t^s\rangle - \langle\psi_t^s,\mathcal{L}^*\mu_t^s\rangle - s\langle\psi_t^s,\partial_s(\mathcal{L}^*\mu_t^s)\rangle\end{aligned}$$

Also by (57) we have

$$\partial_s \operatorname{Ent}_\pi(f_t^s) = \langle\mathcal{L}\psi_t^s,\mu_t^s\rangle = \langle\psi_t^s,\mathcal{L}^*\mu_t^s\rangle.$$

Therefore,

$$\begin{aligned}\frac{1}{2}\partial_t\langle\nabla\psi_t^s,\nabla\psi_t^s\rangle_{\mu_t^s} + \partial_s \operatorname{Ent}_\pi(f_t^s) &= \frac{s}{2}\langle\mathcal{L}^*\mu_t^s\nabla\psi_t^s,\nabla\psi_t^s\rangle - s\langle\psi_t^s,\partial_s(\mathcal{L}^*\mu_t^s)\rangle \\ &= \frac{s}{2}\langle\mathcal{L}^*\mu_t^s,|\nabla\psi_t^s|^2\rangle - s\langle\psi_t^s,\mathcal{L}^*\partial_s\mu_t^s\rangle \\ &= \frac{s}{2}\langle\mathcal{L}^*\mu_t^s,|\nabla\psi_t^s|^2\rangle + s\langle\mathcal{L}\psi_t^s,\operatorname{div}(\mu_t^s\nabla\psi_t^s)\rangle \\ &= -sB(\mu_t^s,\psi_t^s).\end{aligned}$$

References

[1] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev Inequalities for Finite Markov Chains, The Annals of Applied Probability 6 (1996) 695-750.

[2] P. Diaconis and M. Shahshahani, Generating a random permutation with random transposition, Z. Wahr. Verw. Geb. 57 (1981) 159-179.

[3] Sharad Goel, Modified logarithmic Sobolev inequalities for some models of random walk, Stochastic Processes and their Applications 114 (2004) 51-79.

[4] Ronald de Wolf, A Brief Introduction to Fourier Analysis on the Boolean Cube, Theory of Computing Library, Graduate Surveys, TCGS 1 (2008) 1-20.

[5] R. O'Donnell, Lecture notes for a course "Analysis of Boolean functions, 2007. Available at `https://www.cs.cmu.edu/~odonnell/boolean-analysis/`.

[6] S. G. Bobkov and F. Götze, Exponential integrability and transportation-cost related to logarithmic Sobolev inequalities, Journal of Functional Analysis, vol. 163, pp. 1-28, 1999.

[7] N. Gozlan, A characterization of dimension free concentration in terms of transportation inequalities, Annals of Probability, vol. 37, no. 6, pp. 24802498, 2009.

[8] S. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. J. Math. Pures Appl. (9), 80(7):669-696, 2001.

[9] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities and transportation in- equalities. Probab. Theory Related Fields, 133(3):409436, 2005.

[10] J. Lott and C. Villani. Hamilton-Jacobi semigroup on length spaces and applications. J. Math. Pures Appl. (9), 88(3):219-229, 2007.

[11] J. Maas, Gradient flows of the entropy for finite Markov chains, Journal of Functional Analysis Volume 261, Issue 8, Pages 2250-2292, 2011.

[12] E. Mossel, K. Oleszkiewicz and A. Sen, On Reverse Hypercontractivity, Geometric and Functional Analysis **23**(3), 1062-1097 (2013).

[13] M. Erbar and J. Maas, Ricci Curvature of Finite Markov Chains via Convexity of the Entropy, Archive for Rational Mechanics and Analysis, Volume 206, Issue 3, pp 997-1038, 2012.

[14] N. Gozlan and C. Leonard, Transport Inequalities. A Survey, Markov Processes and Related Fields 16 (2010) 635-736

[15] S. Bobkov, An isoperimetric inequality on the discrete cube and an elementary proof of the isoperimetric inequality in Gauss space, Ann. Probability 25, 206-214 (1997).

[16] A. Boucheron, G. Lugosi, and P. Massart, A sharp concentration inequality with applications, Random Structures Algorithms. 16, 277-29 (2000)

[17] S. Bobkov, I. Gentil, and M. Ledoux, Hypercontractivity of Hamilton-Jacobi Equations, J. Math. Pures Appl. 80, 7 (2001) 669-696

[18] L. Gross, Logarithmic Sobolev inequalities, Am. J. Math. **97**, 1061-1083 (1975).

[19] E. Nelson, A quartic interaction in two dimensions, in Mathematical Theory of Elementary Particles (MIT Press, 1966) 69-73.

[20] B. Simon and R. Hoegh-Krohn, Hypercontractive semigroups and two dimensional self-coupled Bose fields, Journal of Functional Analysis **9**(2), 121-180 (1972).

[21] D. Bakry and M. Émery, Diffusions hypercontractives. (French) [Hypercontractive diffusions] Séminaire de probabilités, XIX, 1983/84, 177-206, Lecture Notes in Math., 1123, Springer, Berlin (1985).

[22] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations **26**, 101-174 (2001).

[23] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. **173**, 361-400 (2000).

[24] E. B. Davies, L. Gross, and B. Simon, Hypercontractivity: A bibliographic review, Ideas and Methods in Quantum and Statistical Physics (Cambridge University Press, 1992) 370-389.

[25] G. A. Margulis, Probabilistic characteristics of graphs with large connectivity, Probl. Peredachi Inf. **10**(2), 101-108 (1974).

[26] K. Marton, A simple proof of the blowing-up lemma, IEEE Trans. on Inform. Th. **32**(3), 445-446 (1986).