Geometric Flows for Applied Mathematicians*

Xiaohui Chen

This version: December 24, 2020

Contents

*Working draft in progress and comments are welcome (email: xhchen@illinois.edu).

1 Introduction

Consider particles randomly move on the lattice $\mathbb{Z}^n$. For $x \in \mathbb{Z}^n$ and $t \in \mathbb{Z}_+$, let $p(x,t)$ be the number of "heat" particles in position x at time t. Macroscopically, $p(x,t)$ is observed as the temperature of the lattice system in position x at time t. Since

$$p(x,t+1) = 2^{-n} \sum_{|y-x|_1=1} p(y,t), \tag{1}$$

the discrete version of the time derivative $\partial_t p$ is given by

$$p(x,t+1) - p(x,t) = 2^{-n} \sum_{|y-x|_1=1} [p(y,t) - p(x,t)]. \tag{2}$$

In particular, in dimension $n = 1$, the right-hand side of the last expression equals to

$$\frac{1}{2} \sum_{|y-x|_1=1} [p(y,t) - p(x,t)] = \frac{1}{2}[(p(x+1,t) - p(x,t)) - (p(x-1,t) - p(x,t))] \quad (3)$$

$$\approx \frac{1}{2}\left[\frac{\partial}{\partial x} p(x+1/2,t) - \frac{\partial}{\partial x} p(x-1/2,t)\right] \approx \frac{\partial^2}{\partial x^2} p(x,t). \quad (4)$$

This gives a discrete analog of the heat equation in 1-dimension:

$$\partial_t p = \frac{\partial^2}{\partial x^2} p(x,t). \quad (5)$$

Since particles can only moves orthogonally on the lattice $\mathbb{Z}^n$, the higher-dimensional heat equation for $(x,t) \in \mathbb{R}^n \times \mathbb{R}_+$ (in an appropriate scaling limit) is given by:

$$\partial_t p = \Delta p, \quad (6)$$

where $\Delta p = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} p = \mathrm{tr}(\nabla^2 p) = \mathrm{div}(\nabla p), \nabla p = (\frac{\partial p}{\partial x_1}, \dots, \frac{\partial p}{\partial x_n})^T$, and $\mathrm{div}(v) = \sum_{i=1}^n \frac{\partial v}{\partial x_i}$ for a vector field $v : \mathbb{R}^n \to \mathbb{R}^n$.

2 Heat equation

2.1 Euclidean gradient flow

For $u : \mathbb{R}^n \to \mathbb{R}$, the *Dirichlet energy* is defined as

$$E(u) = \int |\nabla u|^2. \quad (7)$$

Let $u^s = u + tv, t \in \mathbb{R}, v : \mathbb{R}^n \to \mathbb{R}$ be one-parameter family of u. Note that

$$E(u^t) = \int |\nabla u + s\nabla v|^2 = \int |\nabla u|^2 + t^2 \int |\nabla v|^2 + 2t \int \langle \nabla u, \nabla v \rangle. \quad (8)$$

If v has compact support, then the integration-by-parts (cf. Lemma A.3) gives

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} = 2 \int \langle \nabla u, \nabla v \rangle = -2 \int v \mathrm{div}(\nabla u) = -2 \int v \Delta u. \quad (9)$$

By the Cauchy-Schwarz inequality,

$$\int v \Delta u \leqslant \left(\int v^2\right)^{1/2} \left(\int (\Delta u)^2\right)^{1/2} \quad (10)$$

with equality holds if and only if $v = c\Delta u$ for some constant $c \neq 0$. Since $v = \partial_t u^t$, this means that, up to a multiplicative constant, $u_t = \Delta u$ is the (negative) gradient flow for the energy $E(u)$, i.e., if $u_t = \Delta u$ for $u : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} E(u(\cdot,t)) = -2 \int (\Delta u)^2 \leqslant 0. \quad (11)$$

If further $\int |u(\cdot,t)| < \infty$ for each t, then

$$\partial_t \int u(\cdot,t) = \int \partial_t u(\cdot,t) = \int \Delta u(\cdot,t) = 0, \tag{12}$$

where the last equality follows from Lemma A.3. Thus $\int u(\cdot,t)$ is constant in t, which is the *first law of thermodynamics*.

Suppose Ω is a compact subset in $\mathbb{R}^n$ and $u : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ such that $\partial_t u(x,t) = \Delta u(x,t)$ and $u|_{\partial\Omega} = 0$. Since u vanishes on $\partial\Omega$, we have

$$\int_\Omega |\nabla u|^2 = \int_\Omega \langle \nabla u, \nabla u\rangle = -\int_\Omega u\Delta u \leqslant \left(\int_\Omega u^2\right)^{1/2} \left(\int_\Omega (\Delta u)^2\right)^{1/2}. \tag{13}$$

Theorem 2.1 (Poincaré inequality). If $v : \Omega \to \mathbb{R}$ such that $v|_{\partial\Omega} = 0$, then

$$\int_\Omega v^2 \leqslant c \int_\Omega |\nabla v|^2, \tag{14}$$

where c is a constant depending on Ω.

By the Poincaré inequality in Theorem 2.1, we have

$$\int_\Omega |\nabla u|^2 \leqslant \left(c\int_\Omega \nabla u^2\right)^{1/2} \left(-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}E(u)\right)^{1/2}, \tag{15}$$

where c is a constant depending on Ω. In particular, if Ω is connected, then $c > 0$. Then we have

$$E(u) = \left(\int_\Omega \nabla u^2\right)^{1/2} \leqslant c^{1/2} \left(-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}E(u)\right)^{1/2}, \tag{16}$$

i.e.,

$$E(t) \leqslant -\frac{c}{2}E'(t), \qquad \text{where } E(t) := E(u(\cdot,t)). \tag{17}$$

Now by the Gronwall lemma, we have

$$E(t) \leqslant E(0)e^{-2t/c}, \tag{18}$$

which means that the energy of the heat flow decays exponentially fast in time.

2.2 Parabolic maximum principle

Theorem 2.2 (Parabolic maximum principle). Let $\Omega \subset \mathbb{R}^n$ be compact and $u : \Omega \times [0,T] \to \mathbb{R}$. If $(\partial_t - \Delta)u \leqslant 0$ on $\Omega \times [0,T]$ (i.e., $u(x,t)$ is a sub-solution of the heat equation), then

$$\max_{\Omega\times[0,T]} u(x,t) = \max_{(\partial\Omega\times[0,T])\cup(\Omega\times\{0\})} u(x,t). \tag{19}$$

An immediate application of the parabolic maximum principle is the following gradient estimate on solutions of the heat equation.

Theorem 2.3 (Gradient estimate). Let $u : \Omega \times [0,T] \to \mathbb{R}$ such that $\partial_t u = \Delta u$. Then

$$\max_{\Omega\times[0,T]} |\nabla u|^2 = \max_{(\partial\Omega\times[0,T])\cup(\Omega\times\{0\})} |\nabla u|^2. \tag{20}$$

Proof of Theorem 2.2. First assume $(\partial_t - \Delta)u < 0$. If the theorem fails, then there exists (x_0, t_0) in the interior of $\Omega \times [0, T]$ such that

$$u(x_0, t_0) = \max_{\Omega\times[0,T]} u(x, t). \tag{21}$$

Then the first derivative test gives $\nabla u(x_0, t_0) = 0$ and $\partial_t u(x_0, t_0) \geqslant 0$, and the the second derivative test gives $\Delta u(x_0, t_0) \leqslant 0$. Thus at (x_0, t_0), we have $(\partial_t - \Delta)u(x_0, t_0) \geqslant 0$, which is a contradiction to the assumption.

Next consider the general case $(\partial_t - \Delta)u \leqslant 0$. Let $v_\varepsilon(x, t) = u(x, t) - \varepsilon t$ for $\varepsilon > 0$ and $t \in [0, T]$. Then

$$(\partial_t - \Delta)v_\varepsilon = (\partial_t - \Delta)u - \varepsilon \leqslant -\varepsilon < 0, \tag{22}$$

so we can apply the above maximum principle to get

$$\max_{\Omega\times[0,T]} u \geqslant \max_{(\partial\Omega\times[0,T])\cup(\Omega\times\{0\})} u \geqslant \max_{(\partial\Omega\times[0,T])\cup(\Omega\times\{0\})} v_\varepsilon = \max_{\Omega\times[0,T]} v_\varepsilon = \max_{\Omega\times[0,T]} u - \varepsilon T. \tag{23}$$

Now letting $\varepsilon \downarrow 0$, all inequalities in the last display become equality and this finishes the proof. ∎

Proof of Theorem 2.3. Consider $(\partial_t - \Delta)|\nabla u|^2$. Apply the chain rule to get

$$\partial_t |\nabla u|^2 = \langle 2\nabla u, \nabla u_t \rangle, \qquad \text{where } u_t = \partial_t u = \frac{\mathrm{d}}{\mathrm{d}t} u$$

and

$$\begin{aligned}
\Delta|\nabla u|^2 &= \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} \left[\sum_{j=1}^n \left(\frac{\partial u}{\partial x_j} u \right)^2 \right] \\
&= 2 \sum_{i,j=1}^n \left(\frac{\partial^2}{\partial x_i \partial x_j} u \right)^2 + 2 \sum_{i,j=1}^n \frac{\partial}{\partial x_j} u \cdot \frac{\partial^3}{\partial x_i^2 \partial x_i} u \\
&= 2|\nabla^2 u|_F^2 + 2\langle \nabla u, \nabla \Delta u \rangle,
\end{aligned}$$

where $\nabla^2 u$ is the $n \times n$ Hessian matrix of u and $|\cdot|_F$ denotes the Frobenius norm. Then we have

$$\begin{aligned}
(\partial_t - \Delta)|\nabla u|^2 &= 2\langle \nabla u, \nabla u_t \rangle - 2|\nabla^2 u|_F^2 - 2\langle \nabla u, \nabla \Delta u \rangle \\
&= 2\langle \nabla u, \nabla(u_t - \Delta u) \rangle - 2|\nabla^2 u|_F^2 = -2|\nabla^2 u|_F^2 \leqslant 0.
\end{aligned}$$

Now apply the parabolic maximum principle in Theorem 2.2 to conclude. ∎

2.3 Heat kernels

Definition 2.4 (Heat kernel). Let $\Omega \subset \mathbb{R}^n$. A *heat kernel* is a function $H : \Omega \times \Omega \times \mathbb{R}_+ \to \mathbb{R}$ satisfying the following three properties.

1. *Symmetry and non-negativity:* $H(x, y, t) = H(y, x, t) \geqslant 0$ for all $x, y \in \Omega$.

2. For any fixed $y \in \Omega$,
$$(\partial_t - \Delta_x) H = 0, \tag{24}$$
where Δ_x is the Laplacian with respect to the x variable.

3. *Reproducing property:* for $u_0 : \Omega \to \mathbb{R}$ such that $u_0 \in \mathcal{C}_c^0(\Omega)$ (i.e., continuous function with compact support in Ω), we have
$$\int_\Omega u_0(y) H(x, y, t) \, \mathrm{d}y \to u_0(x), \qquad \text{as } t \downarrow 0. \tag{25}$$

Property (ii) requires the heat kernel H is the fundamental solution of the heat equation. In addition, $(x, t) \mapsto \int u_0(y) H(x, y, t) \, \mathrm{d}y$ solves the heat equation because
$$(\partial_t - \Delta) \int u_0(y) H(x, y, t) \, \mathrm{d}y = \int u_0(y) \, (\partial_t - \Delta_x) H(x, y, t) \, \mathrm{d}y = 0.$$

Corollary 2.5 (Semi-group structure of the heat kernel). Let $u_0 : \mathbb{R}^n \to \mathbb{R}$ such that $u_0 \in \mathcal{C}_c^0$. Then
$$P_t u_0(x) := \int_{\mathbb{R}^n} u_0(y) H(x, y, t) \, \mathrm{d}y, \qquad t \geqslant 0 \tag{26}$$
forms a semi-group of linear operators such that
$$(\partial_t - \Delta) P_t u_0(x) = 0, \tag{27}$$
$$P_t u_0(x) \to u_0(x), \qquad \text{as } t \downarrow 0. \tag{28}$$

2.3.1 Heat kernel on $\mathbb{R}^n$

Let $u_0 : \mathbb{R}^n \to \mathbb{R}$ be a continuous function with compact support (i.e., $u_0 \in \mathcal{C}_c^0$). If u_0 does not grow too fast at infinity, then by the uniqueness the solution of (33) is given by
$$u(x, t) = \int_{\mathbb{R}^n} u_0(y) H(x, y, t) \, \mathrm{d}y, \tag{29}$$
where $H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ is the heat kernel on $\mathbb{R}^n$ is defined as
$$H(x, y, t) = (4\pi t)^{-n/2} \exp\left(-\frac{|x - y|^2}{4t} \right). \tag{30}$$

Lemma 2.6 (Heat kernel on $\mathbb{R}^n$). The function H defined in 30 is the heat kernel on $\mathbb{R}^n$.

Proof of Lemma 2.6. Parts (i) and (ii) in Definition 2.4 are obvious. For part (iii), by the continuity of u_0, we have as $t \downarrow 0$,
$$\int_{\mathbb{R}^n} u_0(y) H(x, y, t) \, \mathrm{d}y = \int_{\mathbb{R}^n} u_0(y) \underbrace{(4\pi t)^{-n/2} \exp\left(-\frac{|x - y|^2}{4t} \right)}_{\to \delta_x(y)} \to u_0(x).$$

■

2.3.2 Heat kernel on $\mathbb{S}^1$

Consider $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$. So $\Delta u = u''$. We can find an orthonormal basis of eigenfunctions $(\phi_i)_{i=1}^{\infty}$ for the Laplacian operator Δ solving the equation

$$\Delta\phi_i + \lambda_i\phi_i = 0. \tag{31}$$

Lemma 2.7 (Heat kernel on $\mathbb{S}^1$). The function

$$H(x,y,t) = \sum_{i=1}^{\infty} e^{-\lambda_i t}\phi_i(x)\phi_i(y) \tag{32}$$

is the heat kernel on $\mathbb{S}^1$.

Proof of Lemma 2.7. Symmetry is obvious $H(x,y,t) = H(y,x,t)$. Next we compute

$$\begin{aligned}
\partial_t H(x,y,t) &= \sum_{i=1}^{\infty}(-\lambda_i)e^{-\lambda_i t}\phi_i(x)\phi_i(y),\\
\Delta_x H(x,y,t) &= \sum_{i=1}^{\infty} e^{-\lambda_i t}\Delta\phi_i(x)\phi_i(y) = \sum_{i=1}^{\infty} e^{-\lambda_i t}(-\lambda_i)\phi_i(x)\phi_i(y).
\end{aligned}$$

Thus $(\partial_t - \Delta_x)H(x,y,t) = 0$ for each fixed $y \in \mathbb{S}$. To check the reproducing property, given the expansion $u_0(y) = \sum_{i=1}^{n} a_i\phi_i(y)$, we have

$$\begin{aligned}
\int u_0(y)H(x,y,t)\,\mathrm{d}y &= \int_{\mathbb{S}^1}\sum_{i=1}^{\infty} u_0(y)e^{-\lambda_i t}\phi_i(x)\phi_i(y)\,\mathrm{d}y\\
&= \sum_{i=1}^{\infty} e^{-\lambda_i t}\phi_i(x)\int_{\mathbb{S}^1} u_0(y)\phi_i(y)\,\mathrm{d}y\\
&= \sum_{i=1}^{\infty} e^{-\lambda_i t}\phi_i(x)a_i.
\end{aligned}$$

As $t \downarrow 0$, $e^{-\lambda_i t} \to 1$ on the compact $\mathbb{S}^1$ and

$$\int_{\mathbb{S}^1} u_0(y)H(x,y,t)\,\mathrm{d}y \to \sum_{i=1}^{\infty}\phi_i(x)a_i = u_0(x).$$

Finally, we check the positivity of H. Suppose there exists (x_0,y_0,t_0) such that $H(x_0,y_0,t_0) < 0$. Then we can a continuous function $u : \mathbb{R}^n \to \mathbb{R}$ with compact support in a neighborhood of y_0 where $H(x_0,\cdot,t_0) < 0$ such that $u \geqslant 0$ and $u \neq 0$. By the parabolic maximum principle in Theorem 2.2, we have

$$U(x,t) := \int u(y)H(x,y,t)\,\mathrm{d}y \geqslant 0.$$

But

$$U(x_0,t_0) = \int \underbrace{u(y)}_{>0}\underbrace{H(x_0,y,t_0)}_{<0}\,\mathrm{d}y < 0.$$

So we get a contradiction and we must have $H(x,y,t) \geqslant 0$. ∎

2.4 Central limit theorem

Let $u_0 : \mathbb{R}^n \to \mathbb{R}$ be a continuous function with compact support (i.e., $u_0 \in \mathcal{C}_c^0$). Consider the initial value problem:

$$\begin{cases} (\partial_t - \Delta)u = 0 \\ u(x,0) = u_0(x) \end{cases} . \tag{33}$$

Suppose u_0 does not grow too fast at infinity. Recall that the uniqueness the solution of (33) is given by $u(x,t) = \int_{\mathbb{R}^n} u_0(y)H(x,y,t)\,\mathrm{d}y$, where H is the heat kernel on $\mathbb{R}^n$ defined in (30). A crude bound

$$|u(x,t)| \leqslant (4\pi t)^{-n/2} \int_{\mathbb{R}^n} |u_0| \to 0 \qquad \text{as } t \to \infty \tag{34}$$

shows that the temperature $u(x,t)$ vanishes in the long-time dynamics. The goal of this subsection is to show that the solution of the heat equation in (33) tends to Gaussian after proper rescaling, i.e., the central limit theorem (CLT) behavior.

Theorem 2.8 (Central limit theorem for heat equation). Let

$$v(x,t) = (4\pi t)^{n/2} u(\sqrt{t}x, t). \tag{35}$$

Then as $t \to \infty$,

$$v(x,t) \to \exp(-|x|^2/4) \int_{\mathbb{R}^n} u_0, \tag{36}$$

or equivalently $u(\sqrt{t}x, t) \to G(x) \int_{\mathbb{R}^n} u_0$, where

$$G(x) = (4\pi)^{-n/2} \exp(-|x|^2/4) \tag{37}$$

is the standard Gaussian in $\mathbb{R}^n$.

Proof of Theorem 2.8. Note that

$$\begin{aligned} v(x,t) &= (4\pi t)^{n/2} \int u_0(y)(4\pi t)^{-n/2} \exp\left(-\frac{|\sqrt{t}x - y|^2}{4t}\right) \mathrm{d}y \\ &= \int u_0(y) \exp\left(-\frac{|x|^2}{4} - \frac{|y|^2}{4t} + \frac{2\sqrt{t}\langle x,y\rangle}{4t}\right) \mathrm{d}y \\ &= \exp\left(-\frac{|x|^2}{4}\right) \int u_0(y) \exp\Big(\underbrace{-\frac{|y|^2}{4t}}_{\to 1} + \underbrace{\frac{\langle x,y\rangle}{2\sqrt{t}}}_{\to 1}\Big) \mathrm{d}y \\ &\to \exp\left(-\frac{|x|^2}{4}\right) \int u_0(y)\,\mathrm{d}y \qquad \text{as } t \to \infty. \end{aligned}$$

■

2.4.1 Functional inequalities

The CLT behavior and monotonicity of the heat equation are powerful tools to provide "dynamical proofs" of some well-known functional inequalities. The general idea is to use *interpolation*. To prove a functional inequality, we run two continuous-time heat equations (i.e., heat flows) with the initial data given by the left-hand side of the inequality. A key step is to find a monotonicity (non-decreasing) of the solutions of the heat equations, which means that the left-hand side of the inequality is smaller than any later time point of the functional. On the other hand, if we run the heat equations for a long enough time, then we get a CLT behavior with property rescaling, modulo a constant that depends on the initial data. Those constants preserve the total energy of the two heat flows which are associated to the right-hand side of the inequality. Thus when we look at the long-time dynamics of the heat flows, then the functional with initial data is always less than the time limit. This would prove the desired inequality.

First, we shall demonstrate the above idea to prove the Hölder inequality.

Theorem 2.9 (Hölder inequality). Let $u, v : \mathbb{R}^n \to \mathbb{R}$ such that $u \in L^p$ and $v \in L^q$ for $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\int |uv| \leqslant \left(\int |u|^p \right)^{1/p} \left(\int |v|^q \right)^{1/q}. \tag{38}$$

Proof of Theorem 2.9. *Step 1.* By approximation, it is enough to assume u, v are continuous functions with compact support, i.e., $u, v \in \mathcal{C}_c^0(\mathbb{R}^n)$. Let $f, g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ solve the following initial value problems of the heat equations respectively:

$$\begin{cases} (\partial_t - \Delta) f = 0 \\ f(x,0) = |u(x)|^p \end{cases} \quad \text{and} \quad \begin{cases} (\partial_t - \Delta) g = 0 \\ g(x,0) = |v(x)|^q \end{cases}.$$

We claim that $f, g \geqslant 0$ because we have by the parabolic maximum principle in Theorem 2.2 that

$$(\partial_t - \Delta)(-f) = 0 \implies \max\{-f(x,t)\} = \max\{-|u(x)|^p\} \leqslant 0.$$

Thus $f(x,t) \geqslant 0$, which means that if we start from non-negative initial data, then the solution of the heat equation remains non-negative.

Step 2. Our next claim is:

$$\partial_t \int f^{1/p} g^{1/q} \geqslant 0, \tag{39}$$

which implies that $\int f^{1/p}g^{1/q}$ is non-decreasing in t. Indeed, direct computation yields

$$\begin{aligned}
\partial_t \int f^{1/p}g(1/q) &= \int p^{-1}f^{1/p-1}f_t g^{1/q} + q^{-1}f^{1/p}g^{1/q-1}g_t \\
&= \int p^{-1}f_t f^{-1/q}g^{1/q} + q^{-1}g_t f^{1/p}g^{-1/p} \\
&= \int p^{-1}(\Delta f)f^{-1/q}g^{1/q} + q^{-1}(\Delta g)f^{1/p}g^{-1/p} \\
&= -p^{-1}\int \langle \nabla f, \nabla(f^{-1/q}g^{1/q})\rangle - q^{-1}\int \langle \nabla g, \nabla(g^{-1/p}f^{1/p})\rangle \\
&= -p^{-1}\int \langle \nabla f, -q^{-1}(\nabla f)f^{-1/q-1}g^{1/q} + q^{-1}(\nabla g)f^{-1/q}g^{1/q-1}\rangle \\
&\quad -q^{-1}\int \langle \nabla g, -p^{-1}(\nabla g)g^{-1/p-1}f^{1/p} + g^{-1/p}(\nabla f)f^{1/p-1}\rangle \\
&= (pq)^{-1}\int |\nabla f|^2 f^{-1/q-1}g^{1/q} + (pq)^{-1}\int |\nabla g|^2 g^{-1/p-1}f^{1/p} \\
&\quad -2(pq)^{-1}\int \langle \nabla f, \nabla g\rangle f^{-1/q}g^{-1/p}.
\end{aligned}$$

Note that

$$\left|\nabla \log \frac{f}{g}\right|^2 = \left|\frac{\nabla f}{f} - \frac{\nabla g}{g}\right|^2 = \frac{|\nabla f|^2}{f^2} + \frac{|\nabla g|^2}{g^2} - \frac{2\langle \nabla f, \nabla g\rangle}{fg}.$$

Combining the last two equations and using $p^{-1} + q^{-1} = 1$, we get

$$\partial_t \int f^{1/p}g^{1/q} = \int \frac{1}{pq}\left|\nabla \log \frac{f}{g}\right|^2 f^{1/p}q^{1/q} \geqslant 0 \tag{40}$$

with equality attained if and only if $u = cv$ for some constant $c \neq 0$. This proves the monotonicity (39).

Step 3. By the CLT in Theorem 2.8, we have

$$t^{n/2}f(\sqrt{t}x,t) \to G(x)\int |u|^p \quad \text{and} \quad t^{n/2}g(\sqrt{t}x,t) \to G(x)\int |v|^q \quad \text{as } t \to \infty,$$

where $G(x) = (4\pi)^{-n/2}\exp(-|x|^2/4)$. Since $\partial_t \int f^{1/p}g^{1/q}$ is non-decreasing in t, we have

$$\int |uv| = \int f(x,0)^{1/p}g(x,0)^{1/q} \leqslant \int f(x,t)^{1/p}g(x,t)^{1/q}, \quad \forall t \geqslant 0.$$

Changing variables $x \mapsto \sqrt{t}x$ for $t \geqslant 0$, it follows that

$$\begin{aligned}
\int f(x,t)^{1/p}g(x,t)^{1/q}\,\mathrm{d}x &= \int f(\sqrt{t}x,t)^{1/p}g(\sqrt{t}x,t)^{1/q}t^{n/2}\,\mathrm{d}x \\
&= \int \left[t^{n/2}f(\sqrt{t}x,t)\right]^{1/p}\left[t^{n/2}g(\sqrt{t}x,t)\right]^{1/q}\mathrm{d}x \\
&\nearrow \int \left[G(x)\int |u|^p\right]^{1/p}\left[G(x)\int |v|^q\right]^{1/q}\mathrm{d}x \quad \text{as } t \to \infty \\
&= (\int |u|^p)^{1/p}(\int |v|^q)^{1/q}\int G(x)\,\mathrm{d}x \\
&= (\int |u|^p)^{1/p}(\int |v|^q)^{1/q}.
\end{aligned}$$

Taking $t \to \infty$, we conclude (38). ■

In fact, we can prove a general functional inequality using the CLT which includes the Hölder inequality in Theorem 2.9 as a special case. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a non-decreasing and concave function, i.e., $F_x, F_y \geqslant 0$ and

$$\nabla^2 F = \begin{pmatrix} F_{x,x} & F_{x,y} \\ F_{y,x} & F_{y,y} \end{pmatrix} \leqslant 0$$

as a matrix inequality. If we take $F(x,y) = x^{1/p} y^{1/q}$ for $x, y \geqslant 0, p^{-1} + q^{-1} = 1$ such that $p, q > 1$, then it is easy to check that F is non-decreasing and concave.

Theorem 2.10 (A monotonicity property). Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a non-decreasing and concave function. If the functions $f, g : \mathbb{R}^n \times (0, \infty) \to \mathbb{R}$ such that $(\partial_t - \Delta) f \geqslant 0$ and $(\partial_t - \Delta) g \geqslant 0$, then the function $t \mapsto \int F(f(y,t), g(y,t))\, \mathrm{d}y$ is non-decreasing.

Now if we take $F(x,y) = x^{1/p} y^{1/q}$,

$$\left\{ \begin{array}{c} (\partial_t - \Delta) f = 0 \\ f(x,0) = |u(x)|^p \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{c} (\partial_t - \Delta) g = 0 \\ g(x,0) = |v(x)|^q \end{array} \right. ,$$

then by Theorem 2.10,

$$\begin{aligned} \int |uv| = \int f(x,0)^{1/p} g(x,0)^{1/q} \;&=\; \int F(f(x,0), g(x,0)) \\ &\leqslant\; \int F(f(x,t), g(x,t)) = \int f(x,t)^{1/p} g(x,t)^{1/q}, \quad \forall t \geqslant 0. \end{aligned}$$

Letting $t \to \infty$ and using the CLT (same as Step 3 in proving Theorem 2.9), we recover the Hölder inequality.

The monotonicity property in Theorem 2.10 can be used to prove other inequalities. Below we give another example.

Lemma 2.11.

$$\int \frac{|uv|}{|u| + |v|} \leqslant \frac{\int |u| \int |v|}{\int |u| + \int |v|}. \tag{41}$$

Proof of Lemma 2.11. It suffices to prove that for $u, v \geqslant 0$,

$$\int \frac{uv}{u+v} \leqslant \frac{\int u \int v}{\int u + \int v}.$$

We shall apply Theorem 2.10 with $F(x,y) = \frac{xy}{x+y}$ for $x, y \geqslant 0$. It is easy to check that F is non-decreasing and concave. Take

$$\left\{ \begin{array}{c} (\partial_t - \Delta) f = 0 \\ f(x,0) = u(x) \geqslant 0 \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{c} (\partial_t - \Delta) g = 0 \\ g(x,0) = v(x) \geqslant 0 \end{array} \right. .$$

By Theorem 2.10,

$$\int F(f,g) = \int \frac{f(y,t) g(y,t)}{f(y,t) + g(y,t)}\, \mathrm{d}y \quad \text{is non-decreasing in } t.$$

Combining this with the CLT (cf. Theorem 2.8), we deduce that

$$
\begin{aligned}
\int \frac{uv}{u+v} \leqslant \int\int \frac{f(y,t)g(y,t)}{f(y,t)+g(y,t)}\,\mathrm{d}y &= \int \frac{t^{n/2}f(\sqrt{t}y,t)\,t^{n/2}g(\sqrt{t}y,t)}{t^{n/2}f(\sqrt{t}y,t)+t^{n/2}g(\sqrt{t}y,t)}\,\mathrm{d}y \\
&\to \int \frac{(G(y)\int u)(G(y)\int v)}{G(y)\int u + G(y)\int v}\,\mathrm{d}y \quad \text{as } t\to\infty \\
&= \frac{\int u \int v}{\int u + \int v}\int G(y)\,\mathrm{d}y \\
&= \frac{\int u \int v}{\int u + \int v}.
\end{aligned}
$$

■

Proof of Theorem 2.10. We need to show that $\partial_t \int F(f,g) \geqslant 0$ for $t>0$. By the chain rule and the assumption that f and g are super-solutions of the heat equation, we have

$$
\begin{aligned}
\partial_t \int F(f,g) &= \int F_x f_t + F_y g_t \\
&\geqslant \int F_x \Delta f + F_y \Delta g \\
&= -\int \langle \nabla F_x, \nabla f\rangle + \langle \nabla F_y, \nabla g\rangle \\
&= -\int F_{x,x}|\nabla f|^2 + F_{x,y}\langle \nabla g, \nabla f\rangle + F_{y,x}\langle \nabla f, \nabla g\rangle + F_{y,y}|\nabla g|^2 \\
&= -\int \mathrm{tr}\left[\underbrace{\begin{pmatrix} F_{x,x} & F_{x,y} \\ F_{y,x} & F_{y,y}\end{pmatrix}}_{=\nabla^2 F\leqslant 0}\underbrace{\begin{pmatrix} |\nabla f|^2 & \langle \nabla f, \nabla g\rangle \\ \langle \nabla g, \nabla f\rangle & |\nabla g|^2\end{pmatrix}}_{\geqslant 0 \text{ by Cauchy-Schwarz}}\right] \\
&\geqslant 0.
\end{aligned}
$$

■

2.5 Drift Laplacian

Let $u:\mathbb{R}^n\to\mathbb{R}$ and $\phi:\mathbb{R}^n\to\mathbb{R}$. Define the *drift Laplacian* operator

$$\mathcal{L}_\phi u = \Delta u - \langle \nabla\phi, \nabla u\rangle. \tag{42}$$

It is clear that $\Delta = \mathcal{L}_{\text{constant}}$. Two other important examples we shall discuss later is $\phi(x) = |x|^2/4$ and $\phi(x) = -|x|^2/4$. Note that

$$\mathcal{L}_\phi u = e^{\phi}(e^{-\phi}\Delta u - e^{-\phi}\langle \nabla\phi, \nabla u\rangle) = e^{\phi}\mathrm{div}(e^{-\phi}\nabla u). \tag{43}$$

Suppose that $\int u^2 e^{-\phi} < \infty$ and $\int |\nabla u|^2 e^{-\phi} < \infty$. Then there is a natural inner product associated to $\mathcal{L}_\phi$:

$$\langle u, v\rangle_\phi = \int uve^{-\phi} \tag{44}$$

for $u, v \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, the class of smooth functions with compact support in $\mathbb{R}^n$. Indeed, $\mathcal{L}_\phi$ is a self-adjoint operator w.r.t. the weighted $L^2(e^{-\phi}\mathrm{d}x)$ because

$$\int u\mathcal{L}_\phi v e^{-\phi} = \int u e^{\phi}\mathrm{div}(e^{-\phi}\nabla v)e^{-\phi} = \int u\mathrm{div}(e^{-\phi}\nabla v) = -\int \langle \nabla u, \nabla v\rangle e^{-\phi} = \int v\mathcal{L}_\phi u e^{-\phi}. \tag{45}$$

The inner product $\langle \cdot, \cdot \rangle_\phi$ gives the spectral structure of the drift Laplacian operator $\mathcal{L}_\phi$. If $\mathcal{L}_\phi u + \lambda u = 0$ for some $u \neq 0$, then

$$\int \lambda u^2 e^{-\phi} = -\int u\mathcal{L}_\phi u e^{-\phi} = \int \langle \nabla u, \nabla v\rangle e^{-\phi} = \int |\nabla u|^2 e^{-\phi}.$$

Thus,

$$\lambda = \frac{\int |\nabla u|^2 e^{-\phi}}{\int u^2 e^{-\phi}} \geqslant 0 \tag{46}$$

and $\lambda = 0$ if and only if u is (non-zero) constant. Moreover, if $\mathcal{L}_\phi u + \lambda u = 0$ and $\mathcal{L}_\phi v + \mu v = 0$ such that $\lambda \neq \mu, \mu \neq 0$, and $u, v \neq 0$, then

$$\begin{aligned}
\langle u, v\rangle_\phi &= \int u v e^{-\phi} = \int u \left(\frac{\mathcal{L}_\phi v}{-\mu}\right) e^{-\phi} = -\frac{1}{\mu}\int u\mathcal{L}_\phi v e^{-\phi} \\
&= -\frac{1}{\mu}\int v\mathcal{L}_\phi u e^{-\phi} = \frac{\lambda}{\mu}\int v u e^{-\phi} = \frac{\lambda}{\mu}\langle v, u\rangle_\phi,
\end{aligned}$$

which implies that

$$\langle u, v\rangle_\phi = \int v u e^{-\phi} = 0.$$

Thus the eigenfunctions u and v associated with different eigenvalues are orthogonal.

2.5.1 Ornstein-Uhlenbeck operator

The *Ornstein-Uhlenbeck operator* is the drift Laplacian operator with $\phi(x) = |x|^2/4$, i.e.,

$$L_{OU} u := \mathcal{L}_{\frac{|x|^2}{4}} u = \Delta u - \langle \frac{x}{2}, \nabla u\rangle. \tag{47}$$

It is useful to note that the OU operator can be obtained by the (usual) Laplacian and the heat equation by scaling.

Lemma 2.12 (Connection between heat equation and parabolic OU operator). Let $u(x,t)$ be the solution of the heat equation $(\partial_t - \Delta)u(x,t) = 0$ for $t \in \mathbb{R}$ and $v(x,s) = u(e^{-s/2}x, -e^{-s})$. Then $(\partial_s - L_{OU})v(x,s) = 0$.

Proof of Lemma 2.12. Denote $u_t(x,t)$ as the partial derivative of $u(x,t)$. By the chain rule, we have

$$\begin{aligned}
\partial_s v(x,s) &= -\frac{1}{2}e^{-s/2}\langle x, \nabla u(e^{-s/2}x, -e^{-s})\rangle + e^{-s}u_t(e^{-s/2}x, -e^{-s}), \\
\nabla v(x,s) &= e^{-s/2}\nabla u(e^{-s/2}x, -e^{-s}), \\
\Delta v(x,s) &= e^{-s}\Delta u(e^{-s/2}x, -e^{-s}).
\end{aligned}$$

Then,

$$\begin{aligned}(\partial_s - L_{OU})v(x,s) &= -\frac{1}{2}e^{-s/2}\langle x, \nabla u(e^{-s/2}x, -e^{-s})\rangle + e^{-s}u_t(e^{-s/2}x, -e^{-s}) \\ &\qquad -e^{-s}\Delta u(e^{-s/2}x, -e^{-s}) - \frac{1}{2}\langle x, e^{-s/2}\nabla u(e^{-s/2}x, -e^{-s})\rangle \\ &= e^{-s}(\partial_t - \Delta)u(e^{-s/2}x, -e^{-s}),\end{aligned}$$

showing the desired equivalence. ■

When $n = 1$, the OU operator L_{OU} is also called the *Hermite operator*, i.e.,

$$L_{OU}u = u'' - \frac{1}{2}xu'. \tag{48}$$

It is easy to check that the eigenvalues of the Hermite operator are multiples of $1/2$ and the eigenfunctions are the Hermite polynomials. For instance, $(0, 1)$ is the smallest eigenvalue and eigenfunction pair because $L_{OU}1 = 0$; $(1/2, x)$ is the second smallest eigenvalue and eigenfunction pair because $L_{OU}x + \frac{x}{2} = 0$; $(1, x^2 - 2)$ is the third smallest eigenvalue and eigenfunction pair because $L_{OU}(x^2-2)+(x^2-2) = 0$, etc. Note that the Hermite polynomials are orthogonal polynomials w.r.t. the standard Gaussian measure $e^{-x^2/4}dx$.

2.5.2 Mehler flow

Recall that in Section 2.5.1 (Lemma 2.12), it is shown that $u(x,t)$ solves the heat equation $(\partial_t - \Delta)u = 0$ if and only if $v(x,s) := u(e^{-s/2}x, -e^{-s})$ solves $(\partial_s - L_{OU})v = 0$. In this section, we provide another scaling (forward in time) for the heat equation.

Suppose $(\partial_t - \Delta)u(x,t) = 0$. Let

$$\tilde{u}(x,t) := t^{n/2}u(\sqrt{t}x, t). \tag{49}$$

By the CLT in Theorem 2.8,

$$\tilde{u}(x,t) \to (4\pi)^{-n/2}\exp(-|x|^2/4)\int u, \quad \text{as } t \to \infty.$$

Now changing the time clock $t = e^s$, we define

$$v(x,s) := \tilde{u}(x, e^s) = e^{ns/2}u(e^{s/2}x, e^s). \tag{50}$$

Lemma 2.13 (Connection between heat equation and Mehler flow). Let $u : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and v be defined in (50). Then $u(x,t)$ solves the solution of the heat equation $(\partial_t - \Delta)u(x,t) = 0$ if and only if $v(x,s)$ solves the equation $(\partial_s - L_M)v(x,s) = 0$, where

$$L_M v := \mathcal{L}_{-\frac{|x|^2}{4}}v + \frac{n}{2}v = \Delta v + \frac{1}{2}\langle x, \nabla v\rangle + \frac{n}{2}v \tag{51}$$

is the *Mehler operator*.

Proof of Lemma 2.13. Denote $u_t(x,t)$ as the partial derivative of $u(x,t)$. By the chain rule, we have

$$\begin{aligned}
\partial_s v &= \frac{n}{2}e^{ns/2}u + e^{ns/2}\left[\frac{1}{2}e^{s/2}\langle x, \nabla u\rangle + e^s u_t\right], \\
\nabla v &= e^{ns/2}e^s \nabla u, \\
\Delta v &= e^{ns/2}e^s \Delta u,
\end{aligned}$$

where $v := v(x,s)$ and $u := u(e^{s/2}x, e^s)$. Then,

$$\begin{aligned}
(\partial_s - L_M)v &= \frac{n}{2}e^{ns/2}u + \frac{1}{2}e^{ns/2}e^{s/2}\langle x, \nabla u\rangle + e^{ns/2}e^s u_t \\
&\qquad -e^{ns/2}e^s \Delta u - \frac{1}{2}\langle x, \nabla u\rangle e^{ns/2}e^s - \frac{n}{2}v \\
&= e^{(n/2+1)s}(\partial_t - \Delta)u,
\end{aligned}$$

showing the desired equivalence. ■

Lemma 2.13 shows that we can derive the Mehler flow from the heat flow by properly scaling over space and (forward in) time. Now consider the fundamental solution of the heat equation on $\mathbb{R}^n$:

$$u(x,t) = (4\pi t)^{-n/2}\exp\left(-\frac{|x|^2}{4t}\right).$$

Changing variable $t = e^s$, we get

$$v(x,s) = e^{ns/2}u(e^{s/2}x, e^s) = (4\pi)^{-n/2}\exp\left(-\frac{|x|^2}{4}\right) = G(x),$$

where $G(x)$ is the standard Gaussian on $\mathbb{R}^n$ that does not depend on s. Thus $L_M G = 0$ since $(\partial_s - L_M)G = 0$ by Lemma 2.13. This means that G is a critical point for the equation $(\partial_s - L_M)v(x,s) = 0$ and G is L_M-harmonic.The equation $(\partial_s - L_M)v = 0$ is sometimes referred as the *Mehler flow*.

Suppose we run the Mehler flow $(\partial_s - L_M)v(x,s) = 0$ from an initial data $v(x,0) = v_0(x)$ with compact support. Then by the CLT in Theorem 2.8, we see that as $t = e^s \to \infty$,

$$v(x,s) = (e^s)^{n/2}u(\sqrt{e^s}x, e^s) \to G(x)\int u_0,$$

implying that the Mehler flow $v(x,s)$ converges to a multiple of Gaussian as $s \to \infty$ (i.e., long-time dynamics). The next result shows that the standard Gaussian, modulo multiplicative constant, is the only L_M-harmonic function.

Theorem 2.14 (Gradient flow structure of Melher flow). Let $f(x,t)$ be a function satisfying $f, |\nabla f| \in L^2(e^{|x|^2/4}\mathrm{d}x)$. The Mehler flow $(\partial_t - L_M)f = 0$ is the (negative) gradient flow of the Mehler energy functional

$$E(f) = \int (|\nabla f|^2 - \frac{n}{2}f)e^{|x|^2/4} \geqslant 0. \tag{52}$$

Moreover, $L_M g = 0$ if and only if $g = ce^{-|x|^2/4}$ for some $c > 0$.

Proof of Theorem 2.14. By Lemma 2.15, $E(f) \geqslant 0$ and $E(f) = 0$ if and only if $f(x) = ce^{-|x|^2/4}$. For a function $f(x,t)$, we also write $E(t) = \int(|\nabla f|^2 - \frac{n}{2}f)e^{|x|^2/4}$. Note that by the chain rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = \int(2\langle\nabla f, \nabla f_t\rangle - nff_t)e^{|x|^2/4}$$

and

$$\begin{aligned}\mathrm{div}(f_t e^{|x|^2/4}\nabla f) &= \langle\nabla f_t, e^{|x|^2/4}\nabla f\rangle + f_t e^{|x|^2/4}\langle\frac{x}{2}, \nabla f\rangle + f_t e^{|x|^2/4}\Delta f \\ &= \langle\nabla f_t, \nabla f\rangle e^{|x|^2/4} + f_t e^{|x|^2/4}\mathcal{L}_{-\frac{|x|^2}{4}}f.\end{aligned}$$

We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}E(t) &= 2\int \mathrm{div}(f_t e^{|x|^2/4}\nabla f) - 2\int f_t e^{|x|^2/4}(\mathcal{L}_{-\frac{|x|^2}{4}}f + \frac{n}{2}f) \\ &= -2\int f_t e^{|x|^2/4}L_M f, \end{aligned} \tag{53}$$

where the last equality follows from the divergence theorem under the assumption that $f, |\nabla f| \in L^2(e^{|x|^2/4}\mathrm{d}x)$. Choosing $f_t = L_M f$, we see that

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = -2\int(L_M f)^2 e^{|x|^2/4} \leqslant 0,$$

which means that $E(t)$ is non-increasing in t along the Mehler flow $(\partial_t - L_M)f = 0$. Moreover, the energy $E(t) = 0$ if and only if $L_M f = 0$. This together with Lemma 2.15 imply that $f = ce^{-|x|^2/4}$ is the only L_M-harmonic function. ■

Lemma 2.15. Let $E(\cdot)$ be the Mehler energy defined in (52). If $v, |\nabla v| \in L^2(e^{|x|^2/4}\mathrm{d}x)$, then $E(v) \geqslant 0$ and $E(v) = 0$ if and only if $v(x) = ce^{-|x|^2/4}$ for some $c > 0$.

Proof of Lemma 2.15. This lemma follows from Lemma 2.16 with $g = e^{-|x|^2/4} > 0, \phi(x) = -|x|^2/4, U(x) = n/2$, and $L = L_M$ such that $L_M g = 0$. ■

Lemma 2.16. Let $\phi : \mathbb{R}^n \to \mathbb{R}$ and $U : \mathbb{R}^n \to \mathbb{R}$. Let $\mathcal{L}_\phi w = \Delta w - \langle\nabla\phi, \nabla w\rangle$ and $Lw = \mathcal{L}_\phi w + Uw$. If $g > 0$, $Lg \leqslant 0$ is a sub-solution of L, and $v, |\nabla v| \in L^2(e^{-\phi}\mathrm{d}x)$, then

$$E(v) = \int(|\nabla v|^2 - v^2 U)e^{-\phi} \geqslant 0 \tag{54}$$

and $E(v) = 0$ if and only if $v = cg$ for some $c > 0$ provided that $v \neq 0$.

Proof of Lemma 2.16. Under the assumption $v, |\nabla v| \in L^2(e^{-\phi}\mathrm{d}x)$, integration-by-parts ensures the uniqueness of the solution to $Lg \leqslant 0$. Since $g > 0$, we can take $w = \log g$. Since $\nabla w = \nabla \log g = \frac{\nabla g}{g}$ and

$$\Delta w = \Delta \log g = \mathrm{div}(\nabla \log g) = \mathrm{div}(\frac{\nabla g}{g}) = \frac{\Delta g}{g} - \frac{|\nabla g|^2}{g^2},$$

we have

$$\begin{aligned}\mathcal{L}_\phi w = \mathcal{L}_\phi \log g &= \Delta \log g - \langle \nabla\phi, \nabla \log g\rangle \\ &= \frac{\Delta g}{g} - \frac{|\nabla g|^2}{g^2} - \langle \nabla \phi, \frac{\nabla g}{g}\rangle \\ &= \frac{\mathcal{L}_\phi g}{g} - \frac{|\nabla g|^2}{g^2} \leqslant -U - |\nabla w|^2\end{aligned}$$

with equality attained if and only if $Lg = 0$. Integrating to get

$$\int v^2 \mathcal{L}_\phi w e^{-\phi} \leqslant -\int v^2 U e^{-\phi} - \int v^2 |\nabla w|^2 e^{-\phi}.$$

Using the inner product structure (45) and recalling $v, |\nabla v| \in L^2(e^{-\phi}\mathrm{d}x)$, we have

$$\int v^2 \mathcal{L}_\phi w e^{-\phi} = -\int \langle \nabla v^2, \nabla w\rangle e^{-\phi} = -2\int v\langle \nabla v, \nabla w\rangle e^{-\phi}.$$

Combining the last two displays, we get

$$\begin{aligned}\int v^2 U e^{-\phi} + \int v^2 |\nabla w|^2 e^{-\phi} &\leqslant 2\int v\langle \nabla v, \nabla w\rangle e^{-\phi} \\ &\leqslant 2\int |v||\nabla w||\nabla v| e^{-\phi} \\ &\leqslant \int v^2 |\nabla w|^2 e^{-\phi} + \int |\nabla v|^2 e^{-\phi},\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality and the third inequality from the elementary inequality $2ab \leqslant a^2 + b^2$. Now we obtain that

$$\int v^2 U e^{-\phi} \leqslant \int |\nabla v|^2 e^{-\phi},$$

i.e., $E(v) = \int (|\nabla v|^2 - v^2 U) e^{-\phi} \geqslant 0$, proving (54). Now tracing the equality case, we see that $Lg = 0$ together with $\nabla v = v\nabla w$ for $v \neq 0$ give $E(v) = 0$. This means that

$$\nabla \log g = \nabla w = \frac{\nabla v}{v} = \nabla \log v,$$

i.e., we need $\nabla(\log g - \log v) = 0$, which holds if and only if $\log v = \log g + c$. Thus $E(v) = 0$ if and only if $v = cg$ for some $c > 0$. ■

2.6 Gradient estimates

In this section, we present several gradient estimates for the (drift) harmonic functions and the heat equation.

2.6.1 L^2 gradient estimates

Recall the drift Laplacian operator $\mathcal{L}_\phi u = \Delta u - \langle \nabla\phi, \nabla u\rangle$. The first estimate for the gradient of L_ϕ-harmonic functions is given in terms of the $L^2(e^{-\phi}\mathrm{d}x)$ norm.

Theorem 2.17 (Reverse Poincaré inequality)*.* If $\mathcal{L}_\phi u = 0$, then

$$\int_{B_r} |\nabla u|^2 e^{-\phi} \leqslant \frac{4}{r^2}\int_{B_{2r}} u^2 e^{-\phi}. \tag{55}$$

Proof of Theorem 2.17. Let $\eta \geqslant 0$ be a cutoff function with compact support in B_{2r}. Using the inner product structure (45) in $L^2(e^{-\phi}\mathrm{d}x)$, we have

$$\begin{aligned} 0 &= \int_{\mathbb{R}^n} \eta^2 u \mathcal{L}_\phi u e^{-\phi} = -\int_{\mathbb{R}^n} \langle \nabla(\eta^2 u), \nabla u\rangle e^{-\phi} \\ &= -2\int_{\mathbb{R}^n} \eta u \langle \nabla\eta, \nabla u\rangle e^{-\phi} - \int_{\mathbb{R}^n} \eta^2 |\nabla u|^2 e^{-\phi}. \end{aligned}$$

Then we have

$$\begin{aligned} \int_{\mathbb{R}^n} \eta^2 |\nabla u|^2 e^{-\phi} &= -2\int_{\mathbb{R}^n} \eta u\langle \nabla\eta, \nabla u\rangle e^{-\phi} \\ &\leqslant 2\int_{\mathbb{R}^n} \eta|u||\nabla\eta||\nabla u| e^{-\phi} \\ &\leqslant \frac{1}{2}\int_{\mathbb{R}^n} \eta^2|\nabla u|^2 e^{-\phi} + 2\int_{\mathbb{R}^n} u^2|\nabla\eta|^2 e^{-\phi}, \end{aligned}$$

where the last inequality follows from the absorbing inequality $2ab \leqslant \varepsilon^{-1}a^2 + \varepsilon b^2$ with $\varepsilon = 2, a = \eta|\nabla u|$, and $b = |u||\nabla\eta|$. Thus we get

$$\int_{\mathbb{R}^n} \eta^2 |\nabla u|^2 e^{-\phi} \leqslant 4\int_{\mathbb{R}^n} u^2 |\nabla\eta|^2 e^{-\phi}.$$

Choose

$$\eta = \begin{cases} 1 & \text{on } B_r \\ \text{linear} & \text{on } B_{2r}\setminus B_r \\ 0 & \text{on } \mathbb{R}^n \setminus B_{2r} \end{cases}. \tag{56}$$

Then $|\nabla\eta| \leqslant r^{-1}$ and

$$\int_{B_r} |\nabla u|^2 e^{-\phi} \leqslant \int_{\mathbb{R}^n} \eta^2|\nabla u|^2 e^{-\phi} \leqslant 4\int_{\mathbb{R}^n} u^2|\nabla\eta|^2 e^{-\phi} \leqslant \frac{4}{r^2}\int_{B_{2r}} u^2 e^{-\phi}.$$

■

2.6.2 Bochner formula

Lemma 2.18 (Bochner formula)*.* Let $u : \mathbb{R}^n \to \mathbb{R}$ such that $u \in \mathcal{C}^3(\mathbb{R}^n)$. Then we have

$$\frac{1}{2}\Delta|\nabla u|^2 = |\nabla^2 u|_F^2 + \langle \nabla\Delta u, \nabla u\rangle, \tag{57}$$

where $\nabla^2 u$ is the Hessian of u and $|\nabla^2 u|_F^2 = \sum_{i,j=1}^n (\frac{\partial^2 u}{\partial x_i \partial x_j})^2$ is the squared Frobenius norm of the matrix $\nabla^2 u$.

Proof of Lemma 2.18. Compute $|\nabla u|^2 = \sum_{i=1}^n (\frac{\partial u}{\partial x_i})^2$ and

$$
\begin{aligned}
\Delta|\nabla u|^2 &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2}\left(\frac{\partial u}{\partial x_i}\right)^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial u}{\partial x_j}\left(2\frac{\partial u}{\partial x_i}\frac{\partial^2 u}{\partial x_i \partial x_j}\right) \\
&= 2\sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial^2 u}{\partial x_i \partial x_j}\right)^2 + 2\sum_{i=1}^n \sum_{j=1}^n \frac{\partial u}{\partial x_i}\frac{\partial^3 u}{\partial x_i \partial^2 x_j} \\
&= 2|\nabla^2 u|_F^2 + 2\sum_{i=1}^n \frac{\partial u}{\partial x_i}\sum_{j=1}^n \frac{\partial u}{\partial x_i}\left(\frac{\partial^2 u}{\partial^2 x_j}\right) \\
&= 2|\nabla^2 u|_F^2 + 2\langle \nabla u, \nabla \Delta u\rangle.
\end{aligned}
$$

■

Similarly, we have the Bochner formula for the drift Laplacian.

Lemma 2.19 (Drift Bochner formula). Let $\phi : \mathbb{R}^n \to \mathbb{R}$ and $\mathcal{L}_\phi u = \Delta u - \langle \nabla\phi, \nabla u\rangle$. If $u \in \mathcal{C}^3(\mathbb{R}^n)$, then we have

$$\frac{1}{2}\mathcal{L}_\phi|\nabla u|^2 = |\nabla^2 u|_F^2 + \langle \nabla\mathcal{L}_\phi u, \nabla u\rangle + \nabla^2\phi(\nabla u, \nabla u), \tag{58}$$

where

$$\nabla^2\phi(\nabla u, \nabla v) = \sum_{i,j=1}^n \frac{\partial^2\phi}{\partial x_i \partial x_j}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} = (\nabla u)^T(\nabla^2\phi)(\nabla v)$$

is a symmetric bilinear form.

Remark 2.20 (Drift Bochner formula on Riemannian manifolds). The drift Bochner formula in Lemma 2.19 can be further generalized to the Riemannian manifold (M, g). For $\phi : M \to \mathbb{R}$, define $\mathcal{L}_\phi u = \Delta u - \langle \nabla\phi, \nabla u\rangle_g$. Then

$$\frac{1}{2}\mathcal{L}_\phi|\nabla u|^2 = |\nabla^2 u|_F^2 + \langle \nabla\mathcal{L}_\phi u, \nabla u\rangle_g + (\nabla^2\phi + \mathrm{Ric}_g)(\nabla u, \nabla u), \tag{59}$$

where Ric_g is the Ricci curvature tensor of (M, g). However, proof of (59) is based on differential geometry, which is very different from the proof of Lemma 2.19 that is presented below.

Proof of Lemma 2.19. By Lemma 2.18, we have

$$
\begin{aligned}
\frac{1}{2}\mathcal{L}_\phi|\nabla u|^2 &= \frac{1}{2}\Delta|\nabla u|^2 - \frac{1}{2}\langle \nabla\phi, \nabla|\nabla u|^2\rangle \\
&= |\nabla^2 u|_F^2 + \langle \nabla\Delta u, \nabla u\rangle - \frac{1}{2}\langle \nabla\phi, \nabla|\nabla u|^2\rangle \\
&= |\nabla^2 u|_F^2 + \langle \nabla\mathcal{L}_\phi u, \nabla u\rangle + \langle \nabla\langle\nabla\phi, \nabla u\rangle, \nabla u\rangle - \frac{1}{2}\langle \nabla\phi, \nabla|\nabla u|^2\rangle.
\end{aligned}
$$

Using the chain rule, we have

$$\nabla_u\langle\nabla\phi,\nabla v\rangle := \langle\nabla\langle\nabla\phi,\nabla v\rangle,\nabla u\rangle = \nabla^2\phi(\nabla v,\nabla u)+\nabla^2 v(\nabla\phi,\nabla u).$$

So we have

$$\begin{aligned}\langle\nabla\langle\nabla\phi,\nabla u\rangle,\nabla u\rangle &= \nabla^2\phi(\nabla u,\nabla u)+\nabla^2 u(\nabla\phi,\nabla u),\\ \langle\nabla\phi,\nabla|\nabla u|^2\rangle &= \nabla_\phi\langle\nabla u,\nabla u\rangle = 2\nabla^2 u(\nabla u,\nabla\phi).\end{aligned}$$

Putting all pieces together, we get

$$\begin{aligned}\frac{1}{2}\mathcal{L}_\phi|\nabla u|^2 &= |\nabla^2 u|_F^2+\langle\nabla\mathcal{L}_\phi u,\nabla u\rangle+\nabla^2\phi(\nabla u,\nabla u)+\nabla^2 u(\nabla\phi,\nabla u)-\nabla^2 u(\nabla u,\nabla\phi)\\ &= |\nabla^2 u|_F^2+\langle\nabla\mathcal{L}_\phi u,\nabla u\rangle+\nabla^2\phi(\nabla u,\nabla u).\end{aligned}$$

■

2.6.3 L^∞ gradient estimate

The Bochner formula is useful for bounding the gradient of solutions to L_ϕ-harmonic functions in terms of the L^∞ norm.

Theorem 2.21 (Cacciopoli inequality: L^∞ version of reverse Poincaré inequality). If $\mathcal{L}_\phi u = 0$ such that $\nabla^2\phi \geqslant 0$ (as a matrix inequality), then

$$\sup_{B_r}|\nabla u|^2 \leqslant C(n,r)\sup_{B_{2r}}|u|^2, \tag{60}$$

where $C(n,r)$ is a dimensional constant that may also depend on r.

Proof of Theorem 2.21. By the drift Bochner formula in Lemma 2.19 and using $\mathcal{L}_\phi u = 0, \nabla^2\phi \geqslant 0$, we have

$$\frac{1}{2}\mathcal{L}_\phi|\nabla u|^2 = |\nabla^2 u|_F^2+\langle\nabla\mathcal{L}_\phi u,\nabla u\rangle+\nabla^2\phi(\nabla u,\nabla u) \geqslant |\nabla^2 u|_F^2.$$

Let $\eta:\mathbb{R}^n\to\mathbb{R}$ be the cutoff function defined in (56) with compact support in B_{2r}. Using

$$\begin{aligned}\mathcal{L}_\phi(uv) &= \Delta(uv)-\langle\nabla\phi,\nabla(uv)\rangle\\ &= (\Delta u)v+u(\Delta v)+2\langle\nabla u,\nabla v\rangle-\langle\nabla\phi,\nabla v\rangle u-\langle\nabla\phi,\nabla u\rangle v\\ &= (\mathcal{L}_\phi u)v+(\mathcal{L}_\phi v)u+2\langle\nabla u,\nabla v\rangle,\end{aligned}$$

we have

$$\frac{1}{2}\mathcal{L}_\phi(\eta^2|\nabla u|^2) = \frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2+\frac{1}{2}(\mathcal{L}_\phi|\nabla u|^2)\eta^2+\langle\nabla\eta^2,\nabla|\nabla u|^2\rangle.$$

Then,

$$\begin{aligned}\frac{1}{2}\mathcal{L}_\phi(\eta^2|\nabla u|^2) &\geqslant \frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2+|\nabla^2 u|_F^2\eta^2+2\eta\langle\nabla\eta,\nabla|\nabla u|^2\rangle\\ &= \frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2+|\nabla^2 u|_F^2\eta^2+2\eta\nabla^2 u(\nabla u,\nabla\eta).\end{aligned}$$

Observe that

$$
\begin{aligned}
|2\eta\nabla^2 u(\nabla u,\nabla\eta)| &= 2\left|\eta\nabla^2 u\left(\frac{\nabla u}{|\nabla u|},\frac{\nabla\eta}{|\nabla\eta|}\right)\right| |\nabla u||\nabla\eta| \\
&\leqslant \eta^2\left|\nabla^2 u\left(\frac{\nabla u}{|\nabla u|},\frac{\nabla\eta}{|\nabla\eta|}\right)\right|^2+|\nabla u|^2|\nabla\eta|^2 \\
&\leqslant \eta^2|\nabla^2 u|_F^2+|\nabla u|^2|\nabla\eta|^2.
\end{aligned}
$$

Combining the last two displays, we get

$$
\frac{1}{2}\mathcal{L}_\phi(\eta^2|\nabla u|^2)\geqslant\frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2-|\nabla u|^2|\nabla\eta|^2.
$$

Let $w=\frac{1}{2}\eta^2|\nabla u|^2+Cu^2$, where $C=\frac{1}{2}\max_{B_{2r}}|\frac{1}{2}(\mathcal{L}_\phi\eta^2)-|\nabla\eta|^2|$ is a constant depending on n and r. Then

$$
\begin{aligned}
\mathcal{L}_\phi w &\geqslant \frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2-|\nabla u|^2|\nabla\eta|^2+C\mathcal{L}_\phi u^2 \\
&= \frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2-|\nabla u|^2|\nabla\eta|^2+2C(u\mathcal{L}_\phi u+|\nabla u|^2) \\
&= \frac{1}{2}(\mathcal{L}_\phi\eta^2)|\nabla u|^2-|\nabla u|^2|\nabla\eta|^2+2C|\nabla u|^2 \\
&\geqslant 0.
\end{aligned}
$$

Since L_ϕ is an elliptic operator (cf. equation (46)), by the maximum principle, w must achieve its maximum at ∂B_{2r}. Thus

$$
\max_{B_{2r}} w=\max_{\partial B_{2r}} w=C\max_{\partial B_{2r}} u^2\leqslant C\max_{B_{2r}} u^2,
$$

where the second equality is due to $\eta=0$ on ∂B_{2r}. Now using the cutoff property of η on B_{2r}, we deduce that

$$
\max_{B_r}|\nabla u|^2\leqslant\max_{B_{2r}} w\leqslant C\max_{B_{2r}} u^2.
$$

■

If we look for positive harmonic functions, then better gradient estimate than Theorem 2.21 can be obtained.

Theorem 2.22 (Gradient estimate for positive harmonic functions). If $u>0$ on $B_{2r}\subset\mathbb{R}^n$ and $\Delta u=0$, then

$$
\sup_{B_r}\frac{|\nabla u|}{u}\leqslant\frac{C(n)}{r}, \tag{61}
$$

where $C(n)$ is a dimensional constant.

On one hand, Theorem 2.22 gives an (ellitpic) Harnack inequality for positive harmonic functions on B_{2r}.

Corollary 2.23 (Harnack inequality). If $u>0$ on $B_{2r}\subset\mathbb{R}^n$ and $\Delta u=0$, then

$$
e^{-2C}\leqslant\frac{u(y)}{u(x)}\leqslant e^{2C},\quad\forall\, x,y\in B_r. \tag{62}
$$

Proof of Corollary 2.23. Let $v = \log u$ and

$$f(s) = v(\frac{y-x}{|y-x|}s + x).$$

Then $f(0) = v(x) = \log u(x)$, $f(|y-x|) = \log u(y)$, and

$$f'(s) = \Big\langle \nabla v(\frac{y-x}{|y-x|}s + x), \frac{y-x}{|y-x|} \Big\rangle.$$

By the Cauchy-Schwarz inequality and Theorem 2.22, we get

$$|f'(s)| \leqslant \left| \Big\langle \nabla v(\frac{y-x}{|y-x|}s + x), \frac{y-x}{|y-x|} \Big\rangle \right| \leqslant \frac{C}{r},$$

where $C := C(n)$ is a dimensional constant. Then the fundamental theorem of calculus yields

$$|f(|y-x|) - f(0)| \leqslant \int_0^{|y-x|} |f'(s)| \, \mathrm{d}s \leqslant \frac{C}{r}|y-x|,$$

which implies for all $x, y \in B_r$,

$$\left| \log \frac{u(y)}{u(x)} \right| \leqslant C \frac{|y-x|}{r} \leqslant 2C.$$

This proves the Harnack inequality (62). ■

On the other hand, if $u > 0$ is an entire harmonic function (i.e., $\Delta u = 0$ on $\mathbb{R}^n$), then u is very rigid – in fact it has to be constant on $\mathbb{R}^n$.

Corollary 2.24 (Liouville theorem)*.* If $u > 0$ on $\mathbb{R}^n$ such that $\Delta u = 0$, then u is constant.

Proof of Corollary 2.24. Put $v = \log u$. From Theorem 2.22, we have

$$\sup_{B_r} |\nabla v| \leqslant \frac{C(n)}{r}$$

for any $r > 0$. Letting $r \to \infty$, we see that $\nabla v = 0$ everywhere, i.e., u is constant. ■

Proof of Theorem 2.22. Let $v = \log u$. We may assume that u is not constant for otherwise the proof is trivial. Note that

$$\nabla v = \frac{\nabla u}{u} \quad \text{and} \quad \Delta v = \frac{\Delta u}{u} - \frac{|\nabla u|^2}{u^2} = -\frac{|\nabla u|^2}{u^2} = -|\nabla v|^2.$$

By the Bochner formula in Lemma 2.18 and the matrix trace inequality in Lemma 2.25,

$$\begin{aligned}
\frac{1}{2}\Delta|\nabla v|^2 = |\nabla^2 v|_F^2 + \langle \nabla \Delta v, \nabla v \rangle &= |\nabla^2 v|_F^2 - \langle \nabla |\nabla v|^2, \nabla v \rangle \\
&\geqslant \frac{|\Delta v|^2}{n} - \langle \nabla |\nabla v|^2, \nabla v \rangle = \frac{|\nabla v|^4}{n} - \langle \nabla |\nabla v|^2, \nabla v \rangle.
\end{aligned}$$

Let $\eta \geqslant 0$ be a cutoff function defined in (56) with compact support in B_{2r}. Then,

$$\begin{aligned}
\Delta(\eta^2|\nabla v|^2) &= \eta^2\Delta|\nabla v|^2 + |\nabla v|^2\Delta\eta^2 + 2\langle\nabla\eta^2, \nabla|\nabla v|^2\rangle \\
&\geqslant 2\eta^2\frac{|\nabla v|^4}{n} - 2\eta^2\langle\nabla|\nabla v|^2, \nabla v\rangle + |\nabla v|^2\Delta\eta^2 + 2\langle\nabla\eta^2, \nabla|\nabla v|^2\rangle.
\end{aligned}$$

Note that $\eta^2|\nabla v|^2$ vanishes on the boundary ∂B_{2r} (because u is not constant), its maximum must be achieved in the interior of B_{2r}, i.e., $\eta^2|\nabla v|^2 > 0$ at the maximum. Thus, at the maximum, we have $\eta > 0$, $\nabla(\eta^2|\nabla v|^2) = 0$, and

$$0 \geqslant \Delta(\eta^2|\nabla v|^2) \geqslant \frac{2\eta^2}{n}|\nabla v|^4 - 2\eta^2\langle\nabla|\nabla v|^2, \nabla v\rangle + |\nabla v|^2\Delta\eta^2 + 4\eta\langle\nabla\eta, \nabla|\nabla v|^2\rangle.$$

Since $\nabla(\eta^2|\nabla v|^2) = 0$ implies $\eta^2\nabla|\nabla v|^2 = -2\eta\nabla\eta|\nabla v|^2$ at the maximum, the last display becomes

$$0 \geqslant \Delta(\eta^2|\nabla v|^2) \geqslant \frac{2\eta^2}{n}|\nabla v|^4 + 4\eta|\nabla v|^2\langle\nabla\eta, \nabla v\rangle + |\nabla v|^2\Delta\eta^2 - 8|\nabla v|^2|\nabla\eta|^2.$$

Dividing $|\nabla v|^2$ on both sides and noting that $\Delta\eta^2 = \operatorname{div}(2\eta\nabla\eta) = 2|\nabla\eta|^2 + 2\eta\Delta\eta = 2|\nabla\eta|^2$ because η is a piecewise linear cutoff function, we get

$$\begin{aligned}
0 &\geqslant \frac{2\eta^2}{n}|\nabla v|^2 + 4\eta\langle\nabla\eta, \nabla v\rangle + \Delta\eta^2 - 8|\nabla\eta|^2 \\
&\geqslant \frac{2\eta^2}{n}|\nabla v|^2 - 4\eta|\nabla\eta||\nabla v| + \Delta\eta^2 - 8|\nabla\eta|^2 \\
&= \frac{2\eta^2}{n}|\nabla v|^2 - 4\eta|\nabla\eta||\nabla v| - 6|\nabla\eta|^2.
\end{aligned}$$

With $a = \eta|\nabla v|$, we can write the last inequality as

$$0 \geqslant \frac{2}{n}a^2 - 4|\nabla\eta|a - 6|\nabla\eta|^2.$$

Solving this quadratic inequality for a and , we get

$$a^2 \leqslant c\frac{|\nabla\eta|^2 + n^{-1}|\nabla\eta|^2}{n^{-2}} \quad \text{at the maximum of } \eta|\nabla v| \text{ on } B_{2r},$$

where c is a universal constant. Then we obtain that

$$\sup_{B_r}\frac{|\nabla u|}{u} = \sup_{B_r}|\nabla v| \leqslant \sup_{B_{2r}}\eta|\nabla v| \leqslant C(n)|\nabla\eta| \leqslant \frac{C(n)}{r}.$$

■

Lemma 2.25 (Matrix trace inequality). Let A be an $n \times n$ matrix. Then,

$$|A|_F^2 \geqslant \frac{\operatorname{tr}^2(A)}{n}. \tag{63}$$

Proof of Lemma 2.25. Inequality (63) follows from

$$\operatorname{tr}(A) = \operatorname{tr}(AI_n) \leqslant \sqrt{\operatorname{tr}(A^TA)}\sqrt{\operatorname{tr}(I_n^2)} = \sqrt{n}|A|_F,$$

where we used the Cauchy-Schwarz inequality. ■

2.6.4 Harnack inequalities

In this section, we derive gradient estimates for the heat equation. The following (parabolic) Harnack inequality allows one to compare the heat equation solution at two different time slices.

Theorem 2.26 (Harnack inequality for heat equation). Suppose $u(x,t) > 0$ solves the heat equation $(\partial_t - \Delta)u = 0$ on $\mathbb{R}^n \times \mathbb{R}_+$. For (x_1, t_1) and (x_2, t_2) such that $t_2 > t_1 > 0$, we have

$$u(x_2,t_2) \geqslant u(x_1,t_1)\left(\frac{t_1}{t_2}\right)^{n/2} \exp\left(-\frac{|x_2-x_1|^2}{4(t_2-t_1)}\right). \tag{64}$$

A key ingredient in proving the Harnack inequality in Theorem 2.26 is the following differential Harnack inequality.

Theorem 2.27 (Differential Harnack inequality for heat equation). Suppose $u : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ satisfying $u(x,t) > 0$ and $(\partial_t - \Delta)u = 0$ on $\mathbb{R}^n \times [0,T]$. Then we have

$$\frac{|\nabla u|^2}{u^2}(x,t) - \frac{u_t}{u}(x,t) \leqslant \frac{n}{2t}. \tag{65}$$

Remark 2.28. The differential Harnack inequality (65) is a sharp global gradient estimate for the heat equation solution. Consider the fundamental solution of the heat equation on $\mathbb{R}^n \times \mathbb{R}_+$, which is given by

$$u(x,t) = (4\pi t)^{-n/2} \exp\left(-\frac{|x|^2}{4t}\right).$$

Then

$$\begin{aligned} v &:= \log u = -\frac{n}{2}\log(4\pi t) - \frac{|x|^2}{4t}, \\ v_t &= -\frac{n}{2t} + \frac{|x|^2}{4t^2}, \quad |\nabla v| = -\frac{x}{2t}, \quad |\nabla v|^2 = \frac{|x|^2}{4t^2}. \end{aligned}$$

So we have

$$\frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} = |\nabla v|^2 - v_t = \frac{|x|^2}{4t^2} + \frac{n}{2t} - \frac{|x|^2}{4t^2} = \frac{n}{2t}.$$

■

Proof of Theorem 2.26. Let $v = \log u$ and

$$f(s) = v(x_2 + \frac{x_1 - x_2}{t_2 - t_1}s, t_2 - s).$$

Then $f(0) = v(x_2, t_2), f(t_2 - t_1) = v(x_1, t_1)$, and

$$f'(s) = \left\langle \nabla v(x_2 + \frac{x_1 - x_2}{t_2 - t_1}s, t_2 - s), \frac{x_1 - x_2}{t_2 - t_1} \right\rangle - v_t(x_2 + \frac{x_1 - x_2}{t_2 - t_1}s, t_2 - s).$$

Using the absorbing inequality $2ab \leqslant \varepsilon^2 + \varepsilon^{-1}b^2$ with $\varepsilon = 2, a = \nabla v, b = \frac{|x_1 - x_2|}{t_2 - t_1}$, we get

$$f'(s) \leqslant |\nabla v|\frac{|x_1 - x_2|}{t_2 - t_1} - v_t \leqslant |\nabla v|^2 + \frac{|x_1 - x_2|^2}{4(t_2 - t_1)^2} - v_t.$$

Applying the differential Harnack inequality in Theorem 2.27 at the spacetime point $(x_2 + \frac{x_1-x_2}{t_2-t_1}s, t_2 - s)$, we have

$$|\nabla v|^2 - v_t = \frac{|\nabla u|^2}{u^2} - \frac{u_t}{u} \leqslant \frac{n}{2}\frac{1}{t_2 - s}.$$

Combining the last two displays, we have

$$f'(s) \leqslant \frac{|x_1 - x_2|^2}{4(t_2 - t_1)^2} + \frac{n}{2}\frac{1}{t_2 - s}.$$

Now using the fundamental theorem of calculus, we get

$$\begin{aligned} f(t_2 - t_1) - f(0) &= \int_0^{t_2-t_1} f'(s)\,\mathrm{d}s \\ &\leqslant \int_0^{t_2-t_1} \left[\frac{|x_1 - x_2|^2}{4(t_2 - t_1)^2} + \frac{n}{2}\frac{1}{t_2 - s}\right] \mathrm{d}s = \frac{|x_1 - x_2|^2}{4(t_2 - t_1)} + \frac{n}{2}\log\left(\frac{t_2}{t_1}\right). \end{aligned}$$

This means that

$$\log\left(\frac{u(x_1, t_1)}{u(x_2, t_2)}\right) \leqslant \log\left[\left(\frac{t_2}{t_1}\right)\exp\left(\frac{|x_1 - x_2|^2}{4(t_2 - t_1)}\right)\right],$$

which is the same as

$$\frac{u(x_2, t_2)}{u(x_1, t_1)} \geqslant \left(\frac{t_1}{t_2}\right)\exp\left(-\frac{|x_1 - x_2|^2}{4(t_2 - t_1)}\right), \qquad \forall\, t_2 > t_1 > 0.$$

■

Proof of Theorem 2.27. Let $v = \log u$. Note that

$$\nabla v = \frac{\nabla u}{u}, \quad \Delta v = \frac{\Delta u}{u} - \frac{|\nabla u|^2}{u^2}, \quad v_t = \frac{u_t}{u},$$

so that

$$(\partial_t - \Delta)v = \frac{u_t}{u} - \frac{\Delta u}{u} + \frac{|\nabla u|^2}{u^2} = \frac{|\nabla u|^2}{u^2} = |\nabla v|^2.$$

Define

$$F(x,t) = t\left[\frac{|\nabla u|^2}{u^2}(x,t) - \frac{u_t}{u}(x,t)\right] = t[|\nabla v|^2 - v_t] = -t|\nabla v|^2.$$

We claim that:

$$(\partial_t - \Delta)F \leqslant \frac{F}{t} + 2\langle\nabla F, \nabla v\rangle - \frac{2}{nt}F^2. \tag{66}$$

Given the claim, our goal is to show $F \leqslant n/2$. Assume that F achieves its maximum on $\mathbb{R}^n \times [0,T]$. Without loss of generality, we may assume the maximum is positive because $F(x,0) = 0$. Then at the maximum, we have

$$\nabla F = 0, \quad \partial_t F \geqslant 0, \quad \Delta F \leqslant 0.$$

Using the claim, we have

$$0 \leqslant (\partial_t - \Delta)F \leqslant \frac{F}{t} - \frac{2}{nt}F^2 \quad \text{at the maximum.}$$

This gives $F \leqslant n/2$. It remains to prove the claim (66). Using the chain rule and the Bochner formula in Lemma 2.18, we compute

$$\begin{aligned}\frac{1}{2}(\partial_t - \Delta)|\nabla v|^2 &= \langle \nabla v, \nabla v_t\rangle - \frac{1}{2}\Delta|\nabla v|^2 \\ &= \langle \nabla v, \nabla v_t\rangle - |\nabla^2 v|_F^2 - \langle \nabla \Delta v, \nabla v\rangle \\ &= -|\nabla^2 v|_F^2 + \langle \nabla(\partial_t - \Delta)v, \nabla v\rangle \\ &= -|\nabla^2 v|_F^2 + \langle \nabla|\nabla v|^2, \nabla v\rangle\end{aligned}$$

and

$$(\partial_t - \Delta)v_t = ((\partial_t - \Delta)v)_t = (|\nabla v|^2)_t = 2\langle \nabla v, \nabla v_t\rangle.$$

Then we have

$$\begin{aligned}(\partial_t - \Delta)F &= (|\nabla v|^2 - v_t) + t(\partial_t - \Delta)(|\nabla v|^2 - v_t) \\ &= \frac{F}{t} + t\left[-2|\nabla^2 v|_F^2 + 2\langle \nabla|\nabla v|^2, \nabla v\rangle - 2\langle \nabla v, \nabla v_t\rangle\right] \\ &= \frac{F}{t} + 2t\left[-|\nabla^2 v|_F^2 + \langle \nabla(|\nabla v|^2 - v_t), \nabla v\rangle\right] \\ &= \frac{F}{t} + 2t\left[-|\nabla^2 v|_F^2 + \langle \nabla \frac{F}{t}, \nabla v\rangle\right] \\ &= \frac{F}{t} - 2t|\nabla^2 v|_F^2 + 2\langle \nabla F, \nabla v\rangle.\end{aligned}$$

Now using the matrix trace inequality in Lemma 2.25, we have

$$|\nabla^2 v|_F^2 \geqslant \frac{|\Delta v|^2}{n} = \frac{1}{n}\left(-\frac{F}{t}\right)^2 = \frac{F^2}{nt^2},$$

where the second equality follows from $\Delta v = \partial_t v - |\nabla v|^2 = -F/t$. Combining the last two displays, we get

$$(\partial_t - \Delta)F \leqslant \frac{F}{t} - \frac{2F^2}{nt} + 2\langle \nabla F, \nabla v\rangle.$$

This proves the claim (66). ■

3 Continuity equation

Let $\Omega \subset \mathbb{R}^n$ be a spatial domain. Consider the *continuity equation* (CE):

$$\partial_t \mu_t + \operatorname{div}(\mu_t \mathbf{v}_t) = 0, \tag{67}$$

where μ_t is a probability measure (typically absolutely continuous with a density) on Ω, $\mathbf{v}_t : \Omega \to \mathbb{R}^n$ is a velocity vector field on Ω, and $\nabla \cdot \mathbf{v}$ is the divergence of a vector field $\mathbf{v}$.

There are several meanings of solving the continuity equation (67). Given the vector field $\mathbf{v}_t$, we can speak of a *classical (or strong) solution* as a partial differential equation (PDE) by thinking $\mu_t(x)$ as a differentiable function of two variables x and t. We can also think of a *distributional solution* by integrating against some "nice" class of test functions on (x, t). If the continuity equation (67) is satisfied, then, for any finite time point $T > 0$, we can integrate with a C^1 function $\psi : \Omega \times [0, T] \to \mathbb{R}$ with bounded support and apply integration-by-parts:

$$\begin{aligned} 0 &= \int_0^T \int_\Omega \psi \partial_t \mu_t + \int_0^T \int_\Omega \psi \mathrm{div}(\mu_t \mathbf{v}_t) && (68) \\ &= -\int_0^T \int_\Omega (\partial_t \psi) \mu_t - \int_0^T \int_\Omega \langle \nabla \psi, \mathbf{v}_t \rangle \mu_t, && (69) \end{aligned}$$

where there is no contribution from the boundary because ψ has compact support so that the divergence theorem works. Here we implicitly assumed that the first law of thermodynamics holds (i.e., mass conservation of μ_t) so that there is no mass escapes at the boundary (if Ω is bounded) or near the infinity (if Ω is not bounded). Compared with the strong solution, the distribution solution (69) does not require differentiability by moving the derivative from $\mu_t(x)$ to $\psi(x, t)$.

Suppose further we can interchange ∂_t and $\int_\Omega$ in the first integral of (68) and we keep the second integral in (69), then we can take a smaller class of test functions only in the spatial domain Ω such that the solution space is larger. Why shouldn't we interchange ∂_t and $\int_\Omega$ in the first integral of (69)? This is because if we take test functions depending only on Ω, then it does not make sense to take the time derivative as in (69), which always equals to zero.

Let $\phi : \Omega \to \mathbb{R}$ be such C^1 test function with bounded support. To make sense of differentiation outside of integration, we obviously need $t \mapsto \int_\Omega \phi \mu_t$ is absolutely continuous in t. In addition, we need the identity

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \phi \mu_t - \int_\Omega \langle \nabla \phi, \mathbf{v}_t \rangle \mu_t = 0 \tag{70}$$

to hold pointwise (in t) such that (69) equals to zero. This motivates the following definition.

Definition 3.1 (Weak solution of the continuity equation). We say that the density μ_t is a *weak solution* in the distribution sense if for any C^1 test function $\phi : \Omega \to \mathbb{R}$ with bounded support, the function $t \mapsto \int_\Omega \phi \mu_t$ is absolutely continuous in t, and for each a.e. t, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \phi \mu_t = \int_\Omega \langle \nabla \phi, \mathbf{v}_t \rangle \mu_t. \tag{71}$$

In these notes, we make (more) sense of dynamic behaviors of the continuity equation, and illustration its links to the classical PDEs, probabilities (such as Markov processes and stochastic differential equations), and the trajectory analysis (such as gradient flows). Specifically, we would like to understand the question that how does the weak solution of the continuity equation in an infinite-dimensional metric space (typically Wasserstein) connect with the classical solution of PDEs in a finite-dimensional space (typically Euclidean space and time)? We start from the (linear) heat equation as a concrete example.

3.1 Metric derivative in Wasserstein space

Let $p \geqslant 1$ and $\mathcal{P}_p(\mathbb{R}^n)$ be the collection of probability measures on $\mathbb{R}^n$ such that the *p-Wasserstein distance* is well-defined, i.e.,

$$\mathcal{P}_p(\mathbb{R}^n) = \Big\{ \mu \in \mathcal{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} |x - x_0|^p \mu(\mathrm{d}x) < \infty \text{ for some } x_0 \in \mathbb{R}^n \Big\}, \tag{72}$$

where $\mathcal{P}(\mathbb{R}^n)$ contains all probability measures on $\mathbb{R}^n$. Let W_p be the p-Wasserstein distance

$$W_p^p(\mu,\nu) = \min\left\{ \int |x-y|^p \gamma(\mathrm{d}x,\mathrm{d}y) : \gamma \in \Gamma \right\}, \tag{73}$$

where Γ is the set of all couplings with marginal distributions μ and ν, i.e., $\int \gamma(\cdot,\mathrm{d}y) = \mu(\cdot)$ and $\int \gamma(\mathrm{d}x,\cdot) = \nu(\cdot)$ for $\gamma \in \Gamma$.

Remark 3.2. The space $(\mathcal{P}_p(\mathbb{R}^n), W_p)$ is a metric space. ■

Definition 3.3 (Absolutely continuous curve). A curve $\omega : [0,1] \to X$, where (X,d) is a metric space is *absolutely continuous* if there exists a $g \in L^1([0,1])$ such that for any $t_0 < t_1$,

$$d(\omega(t_0),\omega(t_1)) \leqslant \int_{t_0}^{t_1} g(\tau)\,\mathrm{d}\tau. \tag{74}$$

Such curves are denoted by $\mathrm{AC}(X)$.

Definition 3.4 (Metric derivative). Let $(\mu_t)_{t>0}$ be an absolutely continuous curve in the Wasserstein (metric) space $(\mathcal{P}_p(\mathbb{R}^n), W_p)$. The *metric derivative* at time t of the curve $t \mapsto \mu_t$ w.r.t. W_p is defined as

$$|\mu'|_p(t) = \lim_{h\to 0} \frac{W_p(\mu_{t+h},\mu_t)}{|h|}. \tag{75}$$

We write $|\mu'|(t) = |\mu'|_2(t)$.

Remark 3.5 (Absolute continuous curves can be reparameterized to be 1-Lipschitz continuous). For any $\omega \in \mathrm{AC}(X)$, it can be reparameterized in time to become Lipschitz continuous. Let $G(t) = \int_0^t g(\tau)\,\mathrm{d}\tau$ and $S(t) = \varepsilon t + G(t)$ for $0 < \varepsilon < 1$. It is easy to see that $S(t)$ is continuous and strictly increasing in $t \in [0,L]$. Define $\tilde{\omega}(t) = w(S^{-1}(t))$ for $t \in [0,L]$. Then for any $0 \leqslant t_1 < t_2 \leqslant L$, we have

$$\begin{aligned}
d(\tilde{\omega}(t_1),\tilde{\omega}(t_2)) &= d(w(S^{-1}(t_1)), w(S^{-1}(t_2))) \\
&\leqslant \int_{S^{-1}(t_1)}^{S^{-1}(t_2)} g(\tau)\,\mathrm{d}\tau = G(S^{-1}(t_2)) - G(S^{-1}(t_1)) \\
&= S(S^{-1}(t_2)) - \varepsilon S^{-1}(t_2) - S(S^{-1}(t_1)) + \varepsilon S^{-1}(t_2) \\
&= |t_2 - t_1| - \varepsilon(S^{-1}(t_2) - S^{-1}(t_1)) \leqslant |t_2 - t_1|.
\end{aligned}$$

Thus $\tilde{\omega}$ is a 1-Lipschitz function in $[0,L]$. Now let $\varepsilon \downarrow 0$, we see that each $\omega \in \mathrm{AC}(X)$ can be reparameterized to be 1-Lipschitz continuous. ■

Theorem 3.6 (Rademacher). If $\omega : [0,1] \to X$ is Lipschitz continuous, then the metric derivative $|\omega'|(t)$ exists for almost everywhere $t \in [0,1]$. In addition, for any $0 \leqslant t < s \leqslant 1$, we have

$$d(\omega(t),\omega(s)) \leqslant \int_t^s |\omega'|(\tau)\,\mathrm{d}\tau. \tag{76}$$

3.2 Heat equation revisited

Recall that the heat equation on $\mathbb{R}^n$ is defined as:

$$(\partial_t - \Delta)u = \partial_t u - \mathrm{div}(\nabla u) = 0, \tag{77}$$

where $u : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ and $u(x, t)$ is a two-variable function of space and time. The fundamental solution of the heat equation (77) is given by

$$u(x, t) = H(x, 0, t), \tag{78}$$

where $H(x, y, t)$ is the heat kernel (sometimes also called Green's function):

$$H(x, y, t) = (4\pi t)^{-n/2} \exp\Big(-\frac{|x-y|^2}{4t}\Big), \quad y \in \mathbb{R}^n,\ t > 0. \tag{79}$$

The classical solution u is just a (positive) function of two variables x and t. One can think of $H(x, y, t)$ is the transition density from x to y in time $2t$.

Let $\Omega = \mathbb{R}^n$ and $\mu_t(x) = u(x, t)$. Clearly $\mu_t > 0$ is the probability density of a Gaussian distribution $N(y, 2tI_n)$ and the continuity equation (67) reads

$$\partial_t \mu_t - \operatorname{div}(\mu_t \frac{\nabla \mu_t}{\mu_t}) = 0. \tag{80}$$

In this case, the velocity vector field is given by

$$\mathbf{v}_t(x) = -\frac{\nabla \mu_t}{\mu_t}(x) = -\frac{\nabla u}{u}(x, t), \tag{81}$$

where the last equality is justified by the equivalence of weak solution and the classical PDE solution (because $u(\cdot, \cdot)$ is Lipschitz continuous and $\mathbf{v}_t(\cdot)$ is Lipschitz). Since

$$\nabla u = (4\pi t)^{-n/2} \exp\Big(-\frac{|x-y|^2}{4t}\Big)\Big(-\frac{x}{2t}\Big) = -u\Big(\frac{I_n}{2t}\Big)x = -u\mathbf{v}_t(x), \tag{82}$$

the velocity vector field $\mathbf{v}_t : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map that can be represented by an $n \times n$ matrix:

$$\mathbf{v}_t = \frac{I_n}{2t}. \tag{83}$$

Thus in the heat equation, the velocity vector field $\mathbf{v}_t = (2t)^{-1}I_n$ does not depend on the location x (i.e., location-free) and it dies off as $t \to \infty$. The vanishing velocity means that the particles moving according to the heat equation will eventually converge to an equilibrium distribution given by a harmonic function $\Delta u = 0$. If the boundary value is imposed (either Dirichlet problem or Neumann problem) or the heat growth at infinity is not too fast, then the solution of the harmonic function is unique.

We can compute the metric derivative of the fundamental solution curve of the heat equation. Note that μ_t is the Gaussian density $N(y, 2tI_n)$ for each $t > 0$.

Lemma 3.7 (Wasserstein distance between two Gaussians, cf. Remark 2.31 in [11]). Let $\mu_1 = N(m_1, \Sigma_1)$ and $\mu_2 = N(m_2, \Sigma_2)$. Then the optimal transport map (i.e., the Monge map) from μ_1 to μ_2 is given by

$$T(x) = m_2 + A(x - m_1), \tag{84}$$

where

$$A = \Sigma_1^{-1/2}(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\Sigma_1^{-1/2}, \tag{85}$$

and the squared 2-Wasserstein distance between μ_1 and μ_2 equals to

$$W_2^2(\mu_1,\mu_2) = |m_1 - m_2|^2 + \operatorname{tr}\left\{\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\right\}, \tag{86}$$

where the last term is the *Bures distance* on positive-semidefinite matrices. In particular, if $\Sigma_1\Sigma_2 = \Sigma_2\Sigma_1$, then

$$W_2^2(\mu_1,\mu_2) = |m_1 - m_2|^2 + |\Sigma_1^{1/2} - \Sigma_2^{1/2}|_F^2. \tag{87}$$

In the Gaussian case, we have

$$W_2^2(\mu_{t+h},\mu_t) = |\sqrt{2(t+h)}I_n - \sqrt{2t}I_n|_F^2 = 2n(\sqrt{t+h} - \sqrt{t})^2, \tag{88}$$

which gives a formulas for $|\mu'|(t)$:

$$|\mu'|(t) = \lim_{h\to 0}\sqrt{2n}\left|\frac{\sqrt{t+h} - \sqrt{t}}{h}\right| = \sqrt{2n}\,\frac{1}{2\sqrt{t}} = \sqrt{\frac{n}{2t}}. \tag{89}$$

On the other hand, recalling (83), we get

$$\|\mathbf{v}_t\|_{L^2(\mu_t)}^2 := \int |\mathbf{v}_t(x)|^2\mu_t(\mathrm{d}x) = \frac{1}{4t^2}\int |x|^2\mu_t(\mathrm{d}x) = \frac{\operatorname{tr}(2tI_n)}{4t^2} = \frac{n}{2t}. \tag{90}$$

This implies that

$$|\mu'|(t) = \|\mathbf{v}_t\|_{L^2(\mu_t)} = \sqrt{\frac{n}{2t}}. \tag{91}$$

Equivalence in (91) is a much more general fact.

Theorem 3.8 (Equivalence between metric derivative and velocity vector field)*.* Let $p > 1$ and $\Omega \subset \mathbb{R}^n$ is compact.
Part 1. If $(\mu_t)_{t\in[0,1]}$ is an AC curve in $W_p(\Omega)$, then for any $t \in [0,1]$ a.e., there is a velocity vector field $\mathbf{v}_t \in L^p(\mu_t;\Omega)$ such that:

1. μ_t is a weak solution of the continuity equation $\partial_t\mu_t + \operatorname{div}(\mu_t\mathbf{v}_t) = 0$ in the sense of distribution;

2. for a.e. t, we have $\|\mathbf{v}_t\|_{L^p(\mu_t)} \leqslant |\mu'|_p(t)$, where $\|\mathbf{v}_t\|_{L^p(\mu_t)} = (\int_\Omega |\mathbf{v}_t|^p\,\mathrm{d}\mu_t)^{1/p}$.

Part 2. Conversely, if $(\mu_t)_{t\in[0,1]}$ are measures in $\mathcal{P}_p(\Omega)$ and $\mathbf{v}_t \in L^p(\mu_t;\Omega)$ for each t such that $\int_0^1 \|\mathbf{v}_t\|_{L^p(\mu_t)}\mathrm{d}t < \infty$ solve the continuity equation $\partial_t\mu_t + \operatorname{div}(\mu_t\mathbf{v}_t) = 0$, then

1. $(\mu_t)_{t\in[0,1]}$ is an AC curve in $W_p(\Omega)$;

2. for a.e. t, we have $|\mu'|_p(t) \leqslant \|\mathbf{v}_t\|_{L^p(\mu_t)}$.

Corollary 3.9*.* Let $p > 1$. If $(\mu_t)_{t\in[0,1]}$ is an AC curve in W_p, then the velocity vector field given in part 1 of Theorem 3.8 must satisfy

$$|\mu'|_p(t) = \|\mathbf{v}_t\|_{L^p(\mu_t)}. \tag{92}$$

Corollary 3.9 is an immediate consequence by applying Theorem 3.8. Alternatively we also provide below a simple and direct heuristic optimal transport argument for Corollary 3.9. Since $(\mu_t)_{t\in[0,1]}$ is AC curve in W_p, there is an optimal transport map $T : \Omega \to \Omega$, moving mass from μ_t to μ_{t+h} that minimizes the cost functional

$$\int |x-y|^p \,\mathrm{d}\gamma \quad \text{subject to } \gamma(\cdot, \mathrm{d}y) = \mu_t(\cdot), \gamma(\mathrm{d}x, \cdot) = \mu_{t+h}(\cdot).$$

This implies that

$$W_p^p(\mu_t, \mu_{t+h}) = \int |x - T(x)|^p \,\mathrm{d}\mu_t(x) = \int |(T - \mathrm{id})(x)|^p \,\mathrm{d}\mu_t(x),$$

where $T - \mathrm{id}$ is the displacement map. Then the discretized velocity of mass movement at location x and time t is given by

$$\mathbf{v}_t(x) = \frac{T(x) - x}{h}.$$

Combining the last displays, we get

$$\|\mathbf{v}_t\|^p_{L^p(\mu_t)} = \int |\mathbf{v}_t(x)|^p \,\mathrm{d}\mu_t(x) = \int \left|\frac{T(x)-x}{h}\right|^p \,\mathrm{d}\mu_t(x) = \frac{W_p^p(\mu_t, \mu_{t+h})}{|h|^p}.$$

Letting $h \to 0$, we have

$$|\mu'|(t) = \lim_{h\to 0} \frac{W_p(\mu_t, \mu_{t+h})}{|h|} = \|\mathbf{v}_t\|_{L^p(\mu_t)}.$$

Moreover, the above argument suggests consider the following dynamics for the particle (or mass) trajectory in $\mathbb{R}^n$ (i.e., the Lagrangian coordinates to "follow" the particle along the flow):

$$\begin{cases} y_x'(t) = \mathbf{v}_t(y_x(t)), \\ y_x(0) = x, \end{cases} \tag{93}$$

where $y_x(t)$ is the time t position of the particle initially at $y_x(0) = x$, i.e., it is the *trajectory* of the particle starting from x. For the heat equation, (93) reads:

$$y_x'(t) = (2t)^{-1} y_x(t) \tag{94}$$

with the initial datum $y_x(0) = x$. This is a first-order linear (homogeneous) ordinary differential equation with initial value problem.

Let μ_t be the evolution of the induced probability distributions of the mass by the trajectory dynamics (93). We need to check that $(\mathbf{v}_t, \mu_t)$ solves the continuity equation (67) in the weak sense (cf. Definition 3.1).

Let $Y_t : \Omega \to \Omega$ be the *flow* of the vector field $\mathbf{v}_t$ on Ω defined through $Y_t(x) = y_x(t)$ in (93). Note that Y_t is indeed a flow on Ω since $Y_0(x) = y_x(0) = x$ and $Y_t(Y_s(x)) = Y_t(y_x(s)) = y_{y_x(s)}(t) = y_x(s+t) = Y_{s+t}(x)$ for any $s, t \geqslant 0$, so that Y_t on Ω is a group action of additive group on $\mathbb{R}_+ = [0, \infty)$.

Let $\mu_t = (Y_t)_\sharp \mu_0$ be the pushforward measure of μ_0 by Y_t, i.e., for any $B \subset \Omega$ measurable,

$$\mu_t(B) := ((Y_t)_\sharp \mu_0)(B) = \mu_0(Y_t^{-1}(B)),$$

or equivalently, for any measurable function ϕ on Ω, the change of variables formula holds:

$$\int \phi \, \mathrm{d}(Y_t)_\sharp \mu_0 = \int (\phi \circ Y_t) \, \mathrm{d}\mu_0.$$

Then $(\mu_t, \mathbf{v}_t)$ is a weak solution of the continuity equation $\partial_t \mu_t + \mathrm{div}(\mu_t \mathbf{v}_t) = 0$ in (67) because for any C^1 test function $\phi : \Omega \to \mathbb{R}$ with bounded support,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int \phi \, \mathrm{d}\mu_t &= \frac{\mathrm{d}}{\mathrm{d}t} \int \phi \, \mathrm{d}(Y_t)_\sharp \mu_0 = \frac{\mathrm{d}}{\mathrm{d}t} \int (\phi \circ Y_t) \, \mathrm{d}\mu_0 = \frac{\mathrm{d}}{\mathrm{d}t} \int \phi(y_x(t)) \, \mathrm{d}\mu_0(x) \\
&= \int \langle \nabla \phi(y_x(t)), y_x'(t) \rangle \, \mathrm{d}\mu_0(x) = \int \langle \nabla \phi(y_x(t)), \mathbf{v}_t(y_x(t)) \rangle \, \mathrm{d}\mu_0(x) \\
&= \int \langle \nabla \phi(Y_t), \mathbf{v}_t(Y_t) \rangle \, \mathrm{d}\mu_0 = \int \langle \nabla \phi, \mathbf{v}_t \rangle \, \mathrm{d}(Y_t)_\sharp \mu_0 = \int \langle \nabla \phi, \mathbf{v}_t \rangle \, \mathrm{d}\mu_t.
\end{aligned}$$

3.3 Wasserstein gradient flow

We have seen from Section 2.1 that the heat equation is the (negative) gradient flow of moving particles in $\mathbb{R}^n$. In this section, we show that the heat equation can also viewed as a gradient flow of the entropy in the Wasserstein space of probability measures. To do this, we need first to make sense what do we mean by a "derivative" of the entropy (or more general functionals) in terms of a density in the infinite-dimensional Wasserstein (or metric) space.

3.3.1 First variation

Definition 3.10 (First variation, Chapter 7 in [12]). Let ρ be a density on Ω. Given a functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$, we call $\frac{\delta F}{\delta \rho}(\rho) : \mathcal{P}(\Omega) \to \mathbb{R}$, if it exists, the unique (up to additive constants) measurable function such that

$$\frac{\mathrm{d}}{\mathrm{d}h}\Big|_{h=0} F(\rho + h\chi) = \lim_{h \to 0} \Big| \frac{F(\rho + h\chi) - F(\rho)}{h} \Big| = \int \frac{\delta F}{\delta \rho}(\rho) \, \mathrm{d}\chi \tag{95}$$

holds for every *mean-zero* perturbation density χ such that $\rho + h\chi \in \mathcal{P}(\Omega)$ for all small enough h. The function $\frac{\delta F}{\delta \rho}(\rho) : \Omega \to \mathbb{R}$ is the *first variation* of F at ρ.

In $\mathbb{R}^n$, the first variation behaves like the directional derivatives projected to all possible directions $\chi \in \mathbb{R}^n$. For example, take $F(x) = \frac{1}{2}|x|^2$ for $x \in \mathbb{R}^n$. Then $\nabla F(x) = x$ and for any $y \in \mathbb{R}^n$,

$$\lim_{h \to 0} \Big| \frac{|x + hy|^2 - |x|^2}{2h} \Big| = \lim_{h \to 0} \Big| \langle x, y \rangle + \frac{h}{2}|y|^2 \Big| = \langle x, y \rangle = \langle \nabla F(x), y \rangle. \tag{96}$$

Comparing (95) with (96), we may interpret

$$\int \frac{\delta F}{\delta \rho}(\rho) \, \mathrm{d}\chi = \Big\langle \frac{\delta F}{\delta \rho}(\rho), \chi \Big\rangle \tag{97}$$

as an inner product in a Hilbert space. Thus first variation is an infinite-dimensional analog of the gradient in the finite-dimensional Euclidean space. Thus $\frac{\delta F}{\delta \rho}(\rho)$ can be viewed as a gradient in $\mathcal{P}(\Omega)$.

Note that first variation is defined only up to additive constants since $\int c \, \mathrm{d}\chi = 0$ for any constant $c \in \mathbb{R}$. Below are two important functionals that will be extremely useful in studying heat equation, or more generally the Fokker-Planck equation in Section **??**.

Example 3.11 ("Generalized" entropy). Let $f : \mathbb{R} \to \mathbb{R}$ be a convex superlinear function and

$$F(\rho) = \int f(\rho(x))\,\mathrm{d}x = \int f \circ \rho. \tag{98}$$

Clearly

$$\frac{F(\rho + h\chi) - F(\rho)}{h} = \int \frac{f(\rho(x) + h\chi(x)) - f(\rho(x))}{h\chi(x)} \chi(x)\,\mathrm{d}x.$$

Letting $h \to 0$, we get

$$\frac{\mathrm{d}}{\mathrm{d}h}\Big|_{h=0} F(\rho + h\chi) = \int f'(\rho(x))\chi(x)\,\mathrm{d}x = \int f'(\rho)\,\mathrm{d}\chi, \tag{99}$$

which implies that

$$\frac{\delta F}{\delta \rho}(\rho) = f'(\rho). \tag{100}$$

For the special case where $f(\rho) = \rho \log \rho$ is the entropy, it first variation at ρ is given by

$$\frac{\delta F}{\delta \rho}(\rho) = 1 + \log \rho. \tag{101}$$

Example 3.12 (Potential). Let $V : \Omega \to \mathbb{R}$ be a potential function and the energy functional

$$F(\rho) = \int V\,\mathrm{d}\rho = \int V(x)\rho(x)\,\mathrm{d}x. \tag{102}$$

Compute

$$\frac{F(\rho + h\chi) - F(\rho)}{h} = \int V(x) \frac{\rho(x) + h\chi(x) - \rho(x)}{h}\,\mathrm{d}x = \int V(x)\chi(x)\,\mathrm{d}x = \int V\,\mathrm{d}\chi.$$

Thus we see that

$$\frac{\delta F}{\delta \rho}(\rho) = V \tag{103}$$

is a constant function that does not depend on ρ.

Intuitively, the first variation of a functional F (either entropy or energy) at ρ is the rate of change in distribution for moving the particles on Ω from ρ that minimizes the entropy/energy.

3.3.2 Minimizing movement scheme

Given a functional $F : \mathcal{P}(\Omega) \to \mathbb{R}$, the *minimizing movement scheme* introduced by Jordan-Kinderlehrer-Otto [6] (sometimes also called the JKO scheme) is a time-discretized version of gradient flows that solves a sequence of iterated minimization problems (in the context of $(\mathcal{P}(\Omega), W_2)$):

$$\rho_{k+1}^h = \mathrm{argmin}_{\rho \in \mathcal{P}(\Omega)} F(\rho) + \frac{W_2^2(\rho, \rho_k^h)}{2h}. \tag{104}$$

By strong duality,

$$W_2^2(\rho, \rho_k^h) = \max_{\varphi \in \Phi_c(\Omega)} \int_\Omega \varphi\,\mathrm{d}\rho + \int_\Omega \varphi^c\,\mathrm{d}\rho_k^h, \tag{105}$$

where $\varphi^c(y) = \inf_{x\in\Omega}\{|x-y|^2 - \varphi(x)\}$ is the c-transform[1]. The functions φ realizing the maximum on the right hand side of (105) is called the *Kantorovich potentials* for the transport from ρ to ρ_k^h. For each ρ_k^h, $W_2^2(\rho, \rho_k^h)$ is convex in ρ since it is a supremum of linear functionals in ρ.

Now differentiating (104) w.r.t. ρ, the first-order optimality condition is implied by

$$\frac{\delta F}{\delta \rho}(\rho_{k+1}^h) + \frac{\varphi}{h} = \text{constant}, \tag{106}$$

where φ now is the Kantorovich potential for the transport from ρ_{k+1}^h to ρ_k^h (not in the reversed direction). Here we implicitly assumed the uniqueness of c-concave Kantorovich potential.

From Brenier's polarization theorem, we know that optimal map $\widetilde{T}$ from μ_{t+h} and μ_t to and $\varphi : \Omega \to \mathbb{R}$ are linked through

$$\widetilde{T}(x) = x - \nabla\varphi(x). \tag{107}$$

So the velocity vector from time $t+h$ to t (note the reverse direction!) is given by

$$\widetilde{\mathbf{v}}_t = \frac{\widetilde{T}(x) - x}{h} = -\frac{\nabla\varphi(x)}{h} = \nabla\Big(\frac{\delta F}{\delta \rho}(\rho_{k+1}^h)\Big)(x). \tag{108}$$

Reverting the time direction ($\mathbf{v}_t = -\widetilde{\mathbf{v}}_t$) and letting $h \to 0$ (using the continuity of ρ), we get

$$\mathbf{v}_t(x) = -\nabla\Big(\frac{\delta F}{\delta \rho}(\rho)\Big)(x) \tag{109}$$

and we get the following continuity equation

$$\partial_t \rho - \operatorname{div}\Big(\rho\,\nabla\Big(\frac{\delta F}{\delta \rho}(\rho)\Big)\Big) = 0 \tag{110}$$

in the Wasserstein space of measures.

If we choose the entropy functional $F(\rho) = \int f(\rho(x))\,\mathrm{d}x$ with $f(\rho) = \rho\log\rho$, then

$$\frac{\delta F}{\delta \rho}(\rho) = 1 + \log\rho, \quad \text{and} \quad \nabla\Big(\frac{\delta F}{\delta \rho}(\rho)\Big) = \frac{\nabla\rho}{\rho} = \nabla\log\rho, \tag{111}$$

so that the continuity equation in (110) reads

$$\partial_t \rho - \operatorname{div}(\rho\,\nabla\log\rho) = 0. \tag{112}$$

Now recall that we can rewrite the heat equation as:

$$0 = (\partial_t - \Delta)u = \partial_t u - \operatorname{div}(u\frac{\nabla u}{u}) = \partial_t u - \operatorname{div}(u\nabla\log u), \tag{113}$$

where we can think of $\mu_t = u(\cdot, t)$ and $\mathbf{v}_t = \nabla\log u$ in the continuity equation (67). Thus we see that (112) and (113) are really the same continuity equation associated with the heat equation. However, they are viewed as different gradient flows in the spaces $(\mathcal{P}_2(\Omega), W_2)$ and Ω, respectively. In either case, we call it the *heat flow*.

---

[1] Here c stands for a general cost function and the c-transform in general is defined as $\varphi^c(y) = \inf_{x\in\Omega}\{c(x,y) - \varphi(x)\}$. A function φ is said to be c-concave if there exists a χ such that $\varphi = \chi^c$ and $\Phi_c(\Omega)$ is the set of all c-concave functions.

3.4 Fokker-Planck equation

The Fokker-Planck equation is a diffusion with a drift term, i.e., it is a diffusive PDE with the drift Laplacian. As in the heat equation, we can take two alternative perspectives on the continuity equation structure and the gradient flow in the spaces $(\mathcal{P}_2(\Omega), W_2)$ and Ω.

First, if we choose the functional

$$F(\rho) = \int f(\rho(x))\,\mathrm{d}x + \int V(x)\,\mathrm{d}\rho(x) = \int f \circ \rho + \int V\,\mathrm{d}\rho, \tag{114}$$

where $f(\rho) = \rho \log \rho$ is the entropy and $V : \Omega \to \mathbb{R}$ is a potential (independent of ρ), then the first variation of F at ρ is given by

$$\frac{\delta F}{\delta \rho}(\rho) = 1 + \log \rho + V. \tag{115}$$

Thus,

$$\nabla\Big(\frac{\delta F}{\delta \rho}(\rho)\Big) = \nabla \log \rho + \nabla V = \frac{\nabla \rho}{\rho} + \nabla V, \tag{116}$$

and the continuity equation for this entropy+potential functional F (note that F is convex in ρ) becomes

$$0 = \partial_t \rho - \mathrm{div}\Big(\rho\left(\frac{\nabla \rho}{\rho} + \nabla V\right)\Big) = \partial_t \rho - \Delta \rho - \mathrm{div}(\rho \nabla V), \tag{117}$$

or alternatively, we may write

$$\partial_t \rho = \Delta \rho - \langle \nabla \rho, \nabla V \rangle - \rho \Delta V, \tag{118}$$

which is the *Fokker-Planck equation* (FPE). Hence with the *drift Laplacian operator* $\mathcal{L}_V$:

$$\mathcal{L}_V \rho = \Delta \rho + \langle \nabla \rho, \nabla V \rangle = e^V\,\mathrm{div}(e^{-V} \nabla \rho), \tag{119}$$

the Fokker-Planck equation (118) is nothing but the *drift heat equation*:

$$\partial_t \rho - \mathcal{L}_V \rho - \rho \Delta V = 0, \tag{120}$$

which describes the time evolution of density of particles under the influence of drag forces and random forces (such as in Brownian motion or heat diffusion in the pure diffusion case.)

Note that the special case when $V(x) = \frac{1}{4}|x|^2$, then $\nabla V = \frac{x}{2}, \Delta V = \frac{n}{2}$ and the drift heat equation (120) is the Mehler flow:

$$(\partial_t - L_M)u = 0,$$

where L_M is the Mehler operator defined as

$$L_M(\rho) = \mathcal{L}_{\frac{|x|^2}{4}}(\rho) + \frac{n}{2}\rho.$$

Recall we have seen from Section 2.5.2 that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_t(f) = -2\langle f_t, L_M f \rangle_M = -2\mathcal{E}_t(f), \tag{121}$$

where $\langle f, g\rangle_M = \int_{\mathbb{R}^n} f(x)g(x)e^{\frac{|x|^2}{4}}\,\mathrm{d}x$ defines an inner product, we see that L_M is the twice negative gradient flow of the Mehler energy (52) in $\mathbb{R}^n$. Note that (121) implies that the exponential rate of convergence for the Mehler flow (i.e., the Fokker-Planck equation with $V(x) = \frac{1}{4}|x|^2$) in the Euclidean space $\mathbb{R}^n$: for $t \geqslant 0$,

$$\mathcal{E}_t(f) = \mathcal{E}_0(f)e^{-2t}. \tag{122}$$

In Section **??** below, we shall also look at the rate of convergence of the gradient flow of the Fokker-Planck equation (in particular the Mehler flow) to v in the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$.

For a general potential V, from the continuity equation (117), the velocity vector field $\mathbf{v}_t$ is given by

$$\mathbf{v}_t = -\frac{\nabla \rho_t}{\rho_t} - \nabla V. \tag{123}$$

Thus we can write down the gradient flow of $\mathbf{v}_t$ as in (93). In particular, the Mehler flow (with the potential $V(x) = \frac{1}{4}|x|^2$) is given by

$$y_x'(t) = -\Big(\frac{\nabla \rho_t}{\rho_t} + \nabla V\Big)(y_x(t)) = -\Big(\frac{\nabla \rho_t}{\rho_t}\Big)(y_x(t)) - \frac{y_x(t)}{2} \tag{124}$$

with the initial datum $y_x(0) = x$, which again is a first-order linear (homogeneous) ordinary differential equation with initial value problem.

3.5 Langevin diffusion

The Fokker-Planck equation is a PDE that describes the time evolution of density of particles of a diffusion process with a (deterministic) drift and (random) noise, which is governed by the following stochastic differential equation (SDE) (sometimes referred as the *Langevin diffusion*):

$$\mathrm{d}X_t = m(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t, \tag{125}$$

where $m(X_t, t)$ is the drift coefficient vector in $\mathbb{R}^n$ and $\sigma(X_t, t)$ is an $n \times n$ matrix. Here (W_t) is again the standard Brownian motion on $\mathbb{R}^n$. The SDE in (125) is understood in the integral form:

$$X_{t+s} - X_s = \int_s^{s+t} m(X_u, u)\,\mathrm{d}u + \int_s^{s+t} \sigma(X_u, u)\,\mathrm{d}W_u \tag{126}$$

as a sum of an Lebesgue integral and an Itô integral. In this model, both the random drift coefficient vector $m(X_t, t)$ and the random matrix $\sigma(X_t, t)$ are path/state and time dependent.

The (standard) n-dimensional Brownian motion is the special case where $m(X_t, t) = 0$ and $\sigma(X_t, t) = I_n$. For $n = 1$, the density evolution equation of (125) is given by

$$\partial_t \rho(x, t) = -\partial_x(m(x, t)\rho(x, t)) + \partial_x^2(D(x, t)\rho(x, t)), \tag{127}$$

where $D(X_t, t) = \sigma(X_t, t)^2/2$ is the diffusion coefficient. In Section 3.5.1, we show how to convert the SDE of sample paths (125) (i.e., Langevin dynamics) to the evolution PDE of densities (127) (i.e., Fokker-Planck equation). In the one-dimensional heat diffusion case (where there is no drift $m(x, t) = 0$) with $D(X_t, t) = 1$ (i.e., $\sigma(X_t, t) = \sqrt{2}$), the evolution of the density $\rho(x, t)$ is governed by

$$\partial_t \rho - \partial_x^2 \rho = 0, \tag{128}$$

which is just the heat equation on $\mathbb{R}$. Higher dimension analog $\partial_t u - \Delta u = 0$ can also be made by noting that the density evolution equation now becomes

$$\partial_t \rho(x,t) = -\mathrm{div}(m(x,t)\rho(x,t)) + \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}(x,t)\rho(x,t)), \tag{129}$$

where $D(x,t) = \sigma(x,t)\sigma(x,t)^T/2$ is the diffusion tensor.

How does the density evolution equation (129) link to the continuity equation version (117) and (118)? With the diffusion tensor $D(x,t) = I_n$, (129) reads

$$\partial_t \rho = -\mathrm{div}(m\rho) + \Delta\rho. \tag{130}$$

Comparing the last display with the continuity equation (117):

$$\partial_t \rho - \Delta\rho - \mathrm{div}(\rho\,\nabla V) = 0, \tag{131}$$

we see that

$$m = -\nabla V, \tag{132}$$

which means that the mean drift vector m in the Fokker-Planck equation proceeds in the negative gradient direction of minimizing the potential V (and of course subject to the diffusion effect given by the Laplacian Δ). Combining this with the fact that the heat flow proceeds in the negative gradient direction in the entropy, we see that the Fokker-Planck equation is the negative gradient flow of the entropy+potential functional

$$F(\rho) = \underbrace{\int f \circ \rho}_{\text{microscopic behavior}} + \underbrace{\int V \,\mathrm{d}\rho}_{\text{macroscopic behavior}} \tag{133}$$

in the Wasserstein space $(\mathcal{P}_2(\Omega), W_2)$.

We remark that in the special case $m(x,t) = -\frac{x}{2}$ (i.e., $V(x) = \frac{1}{4}|x|^2$) and $D(x,t) = 1$ in the Langevin diffusion (125) is often called the *Ornstein-Uhlenbeck (OU) process* in $\mathbb{R}^n$:

$$\mathrm{d}X_t = -\frac{X_t}{2}\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{134}$$

which is simply a mean-reverting diffusion process. The semi-group operator of the OU process is given by

$$P_t f(x) = \mathbb{E}_{\xi\sim\pi}\left[f\Big(e^{-t}\frac{x}{2} + \sqrt{1-e^{-2t}}\xi\Big)\right], \quad t \geqslant 0, \tag{135}$$

where π is again $\sqrt{2}$-scaled standard Gaussian measure γ on $\mathbb{R}^n$, i.e., $\pi(x) = (4\pi)^{-n/2}\exp(-|x|^2/4)$. The OU semi-group $(P_t)_{t\geqslant 0}$ admits π as stationary measure, where the convergence holds in $L^2(\pi)$: for any bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned}
\|P_t f - \pi f\|^2_{L^2(\pi)} &= \mathbb{E}_{\xi'\sim\pi}\left|\mathbb{E}_{\xi\sim\pi}[f(e^{-t}\frac{\xi'}{2} + \sqrt{1-e^{-2t}}\xi)] - \mathbb{E}_{\xi\sim\pi}[f(\xi)]\right|^2 \\
&\leqslant \mathbb{E}_{\xi'\sim\pi}\,\mathbb{E}_{\xi\sim\pi}\left[f(e^{-t}\frac{\xi'}{2} + \sqrt{1-e^{-2t}}\xi) - f(\xi)\right]^2 && (136) \\
&= \mathbb{E}\left[f(e^{-t}\frac{\xi'}{2} + \sqrt{1-e^{-2t}}\xi) - f(\xi)\right]^2 && (137) \\
&\to 0, \quad \text{as } t \to \infty, && (138)
\end{aligned}$$

where (136) follows from Jensen's inequality, (137) from Fubini's theorem, and (138) from the dominated convergence theorem (since $f \in L^2(\pi)$). In addition, the generator $\mathcal{A}$ of $(P_t)_{t\geqslant 0}$ is given by

$$\mathcal{A}f = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} P_t f = -\frac{1}{2}\langle \nabla f, x\rangle + \Delta f, \tag{139}$$

which is called the Ornstein-Uhlenbeck (OU) operator (also write $\mathcal{A} = L_{OU}$). Comparing (139) with the Mehler operator in (**??**):

$$L_M(\rho) = \mathcal{L}_{\frac{|x|^2}{4}}(\rho) + \frac{n}{2}\rho = \Delta\rho + \frac{1}{2}\langle \nabla\rho, x\rangle + \frac{n}{2}\rho, \tag{140}$$

which gives the forward Fokker-Planck equation (cf. Appendix **??** for more details), the generator in (139) gives the backward Fokker-Planck equation. Integrating-by-parts w.r.t. $\mathrm{d}x$, we conclude that $L_M = \mathcal{A}^*$, which means that the Mehler operator (forward equation) is the adjoint of the generator, i.e., the OU operator (backward equation); that is, we have $L_M = L_{OU}^*$ in $L^2(\mathrm{d}x)$. This holds for a general potential V, not just $V(x) = \frac{1}{4}|x|^2$. (Here we need to be slightly careful on the reference measure: if we consider $L^2(e^{-\frac{|x|^2}{4}}\,\mathrm{d}x) = L^2(\mathrm{d}\pi)$, then $\mathcal{L}_{\frac{|x|^2}{4}} = L_{OU}^*$ in $L^2(\mathrm{d}\pi)$.)

To summarize, given a general Fokker-Planck equation:

$$\begin{aligned} 0 &= \partial_t\rho_t - \Delta\rho_t - \mathrm{div}(\rho_t\,\nabla V) && (141)\\ &= \partial_t\rho_t - \mathcal{L}_V\rho_t - \rho_t\Delta V, && (142)\end{aligned}$$

where (141) is the continuity equation version and (142) is the drifted heat equation version, if we assume it admits a stationary distribution $\pi(x) = \frac{1}{Z}e^{-V(x)}$ on $\mathbb{R}^n$ (cf. Section **??** ahead for more details), then the generator $\mathcal{A}$ of the drift heat diffusion process (i.e., the Langevin diffusion) is given by

$$\mathcal{A} = L_V^*, \tag{143}$$

where

$$L_V = \mathcal{L}_V + \rho\Delta V = \Delta\rho + \langle\nabla\rho, \nabla V\rangle + \Delta V, \tag{144}$$

and the semi-group $(P_t)_{t\geqslant 0}$ is given by

$$P_t f(x) = \mathbb{E}_{\xi\sim\pi}[f(e^{-t}\nabla V(x) + \sqrt{1-e^{-2t}}\xi)], \quad t \geqslant 0. \tag{145}$$

Letting $t \to \infty$, we see that P_t asymptotically converges to the stationary distribution π (in $L^2(\pi)$). Thus from the evolution PDE of probability density functions, we can fully characterize the related SDEs. The reverse direction from SDEs to PDEs can be found in Appendix **??**, where we use Itô's formula to show that a measure solution ρ_t to the continuity equation can be seen as the law at time t of the process $(X_t)_{t>0}$ solution to the SDE. This justifies the equivalence between the PDEs and the SDEs.

3.5.1 Feynman-Kac formula

In this section, we show the equivalence between the Langevin SDE and the Fokker-Planck PDE. We begin with one-dimensional derivation $n = 1$. Consider the Langevin diffusion (or sometimes called the *Itô process*):

$$\mathrm{d}X_t = m(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t, \tag{146}$$

where $(W_t)_{t\geqslant 0}$ is the standard Brownian motion in $\mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function. By Itô's formula (cf. Lemma B.1), we have

$$\mathrm{d}f(X_t) = (m\partial_x f + \frac{1}{2}\sigma^2\partial_x^2 f)\,\mathrm{d}t + \sigma(\partial_x f)\,\mathrm{d}W_t,$$

where $m := m(X_t, t)$ and $\sigma := \sigma(X_t, t)$. Taking expectation on both sides, we get

$$\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}f(X_t)\right] = \mathbb{E}\left[m\partial_x f + \frac{1}{2}\sigma^2\partial_x^2 f\right] + \mathbb{E}\left[\sigma(\partial_x f)\,\frac{\mathrm{d}W_t}{\mathrm{d}t}\right].$$

Since $\sigma(\partial_x f)\,\frac{\mathrm{d}W_t}{\mathrm{d}t}$ is a martingale, the second term on the right-hand side is zero. For small time increment Δt, we can approximate the left-hand side of the last equation by

$$\begin{aligned}
\mathbb{E}\left[\frac{f(X_{t+\Delta t}) - f(X_t)}{\Delta t}\right] &= \int_{\mathbb{R}} \frac{f(x)\rho(x, t+\Delta t) - f(x)\rho(x,t)}{\Delta}\,\mathrm{d}x \\
&= \int_{\mathbb{R}} f(x)\frac{\rho(x, t+\Delta t) - \rho(x,t)}{\Delta}\,\mathrm{d}x \\
&\to \int_{\mathbb{R}} f(x)\partial_t\rho(x,t)\,\mathrm{d}x \qquad \text{as } \Delta t \to 0,
\end{aligned}$$

where $\rho_t := \rho(\cdot, t)$ is the probability density of X_t. Thus we get

$$\int_{\mathbb{R}} f(x)\partial_t\rho(x,t)\,\mathrm{d}x = \int_{\mathbb{R}} m(x,t)\partial_x f(x,t)\rho(x,t)\,\mathrm{d}x + \int_{\mathbb{R}} \frac{1}{2}\sigma^2(x,t)\partial_x^2 f(x,t)\rho(x,t)\,\mathrm{d}x.$$

Now integration-by-parts (without boundary term) on the right-hand side of the last display gives

$$\int f\partial_t\rho_t = -\int f\partial_x(m\rho_t) + \int f\partial_x^2\left(\frac{1}{2}\sigma^2\rho_t\right).$$

Since f is arbitrary, we must have

$$\partial_t\rho_t = -\partial_x(m\rho_t) + \partial_x^2\left(\frac{1}{2}\sigma^2\rho_t\right) \tag{147}$$

holds in the distribution sense. With the multivariate Itô's formula in Lemma B.1, higher-dimensional analog of (147) is given by:

$$\partial_t\rho_t = -\mathrm{div}(m\rho_t) + \sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i\partial x_j}\left(D_{ij}\rho_t\right), \tag{148}$$

where $m := m(X_t, t) \in \mathbb{R}^n$ and $D = (D_{ij})_{i,j=1}^n$ is an $n \times n$ diffusion tensor matrix defined as $D = \frac{1}{2}\sigma(X_t,t)\sigma(X_t,t)^T$.

Example 3.13. Let $V : \mathbb{R}^n \to \mathbb{R}$. Take $m(x,t) = \nabla V(x)$ be the gradient vector of V (that is independent of t) and $\sigma(x,t) = \sqrt{2}I_n$. So $D = I_n$ and

$$X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

Then the Fokker-Planck equation reads

$$\begin{aligned}\partial_t \rho_t &= \operatorname{div}(\rho_t \nabla V) + \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} \rho_t = \operatorname{div}(\rho_t \nabla V) + \Delta \rho_t \\ &= \langle \nabla \rho_t, \nabla V \rangle + \rho_t \Delta V + \Delta \rho_t,\end{aligned}$$

which recovers the Wasserstein gradient flow of the entropy+potential functional $F(\rho) = \int \rho \log \rho + \int V \, \mathrm{d}\rho$.

In fact, we have derived the forward and backward equations for the continuous-time Itô process (146).

Theorem 3.14 (Feynman-Kac formula). Let $f(x,t), t \in [0,T]$ and (X_t) be the Itô process in (146) with $m := m(x,t), \sigma := \sigma(x,t)$. The solution of the following PDE

$$\begin{cases} \partial_t f + m \partial_x f + \frac{1}{2} \sigma^2 \partial_x^2 f = 0 \\ f(x,T) = \psi(x) \qquad\qquad \text{(terminal condition)} \end{cases} \tag{149}$$

is given by

$$f(x,t) = \mathbb{E}[\psi(X_T) | X_t = x] = \mathbb{E}[f(X_T, T) | X_t = x]. \tag{150}$$

Proof of Theorem 3.14. Assume the existence of the solution f. By Itô's formula (cf. Lemma B.1),

$$\mathrm{d} f(X_t, t) = (\partial_t f + m \partial_x f + \frac{1}{2} \sigma^2 \partial_x^2 f) \, \mathrm{d}t + \sigma(\partial_x f) \, \mathrm{d}W_t.$$

Under the PDE constraint $\partial_t f + m \partial_x f + \frac{1}{2} \sigma^2 \partial_x^2 f = 0$, we integrate to get for $T > t$,

$$f(X_T, T) - f(X_t, t) = \int_t^T \sigma(\partial_x f) \, \mathrm{d}W_t.$$

Note that $\int_t^T \sigma(\partial_x f) \, \mathrm{d}W_t$ is a martingale because it is an Itô integral of a martingale. Taking conditional expectation given X_t, we have

$$\mathbb{E}[f(X_T, T) | X_t] - f(X_t, t) = \mathbb{E}\left[\int_t^T \sigma(\partial_x f) \, \mathrm{d}W_t | X_t \right] = 0,$$

i.e., $f(X_t, t) = \mathbb{E}[f(X_T, T) | X_t] = \mathbb{E}[\psi(X_T) | X_t]$. ∎

Now it is convenient to define a differential operator $\mathcal{A}$ via

$$\mathcal{A} f = m \partial_x f + \frac{1}{2} \sigma^2 \partial_x^2 f. \tag{151}$$

The operator (151) gives the backward equation $\partial_t f + \mathcal{A} f = 0$. The adjoint operator $\mathcal{A}^*$ in $L^2 \mathrm{d}x$ is determined by:

$$\langle \mathcal{A} f, \rho \rangle = \langle f, \mathcal{A}^* \rho \rangle, \qquad \forall \, \rho, f \in L^2(\mathrm{d}x),$$

where $\langle \rho, f \rangle = \int \rho f$. Using integration-by-parts and (151), we get

$$\begin{aligned}\int (\mathcal{A} f) \rho &= \int (m \partial_x f + \frac{1}{2} \sigma^2 \partial_x^2 f) \rho \\ &= \int -\partial_x (m \rho) f + \partial_x^2 (\frac{1}{2} \sigma^2 \rho) f =: \int (\mathcal{A}^* \rho) f,\end{aligned}$$

where

$$\mathcal{A}^*\rho = -\partial_x(m\rho) + \partial_x^2(\frac{1}{2}\sigma^2\rho). \tag{152}$$

The adjoint operator $\mathcal{A}^*\rho$ in (152) gives the forward equation:

$$\partial_t\rho_t = \mathcal{A}^*\rho = -\partial_x(m\rho) + \partial_x^2(\frac{1}{2}\sigma^2\rho), \tag{153}$$

which describes the evolution of probability densities of the Itô process (146).

For a multivariate Itô process, the forward operator is give by:

$$\mathcal{A}^*\rho = -\mathrm{div}(m\rho) + \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{1}{2}\sigma\sigma^T\rho\right) \tag{154}$$

and the backward operator is given by:

$$\mathcal{A}f = \langle m, \nabla f\rangle + \left\langle \frac{1}{2}\sigma\sigma^T, \nabla^2 f\right\rangle, \tag{155}$$

where $f, \rho : \mathbb{R}^n \to \mathbb{R}$.

3.5.2 Generator and semi-group

Recall that the fundamental solution of the heat equation $(\partial_t - \Delta)u = 0$ on $\mathbb{R}^n$ is the heat kernel:

$$H(x, y, t) = (4\pi t)^{-n/2} \exp\left(-\frac{|x-y|^2}{4t}\right), \qquad x, y \in \mathbb{R}^n,\ t > 0.$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function and define

$$P_t f(x) = \int_{\mathbb{R}^n} f(z) H(x, z, t)\, \mathrm{d}z, \qquad x \in \mathbb{R}^n,\ t > 0. \tag{156}$$

Then one can check that:

1. $P_{t+s}f = P_t(P_s f) = P_s(P_t f)$ for all $t, s \geqslant 0$;
2. $\lim_{t\downarrow 0} P_t f(x) = f(x)$, i.e., the reproducing property.

These two properties means that $(P_t)_{t\geqslant 0}$ forms a continuous *semi-group* of linear operators on $L^2(\pi)$, where $\pi(x) = (4\pi)^{-n/2}\exp(-|x|^2/4)$ is the standard Gaussian on $\mathbb{R}^n$. The linear operators defined in (156) is called the *heat semi-group.*

Definition 3.15 (Markov process). A continuous-time stochastic process $(X_t)_{t\geqslant 0}$ on $\mathbb{R}^n$ is said to be a (homogeneous) *Markov process* if there exists a semi-group of linear operators $(P_t)_{t\geqslant 0}$ satisfying property 1 and 2, and

$$\mathbb{E}[f(X_{t+s})|X_r, r \leqslant t] = (P_s f)(X_t) \qquad \text{almost everywhere}, \tag{157}$$

for all bounded measurable function $f : \mathbb{R}^n \to \mathbb{R}$ and all $s, t \geqslant 0$.

In Definition 3.15, the semi-group (P_s) specifies the dynamics of (X_t) and it does not depend on t. An important function class in testing the Markov property is the indicator functions $f(x) = \mathbf{1}(x \leqslant y)$. In such as, the Markov property reads

$$\mathbb{P}(X_{t+s} \leqslant y | X_r, r \leqslant t) = (P_s f)(X_t). \tag{158}$$

Definition 3.16 (Generator of Markov process). Let $(X_t)_{t\geqslant 0}$ be a continuous-time Markov process with the semi-group (P_t). The *generator* of (X_t) is defined as

$$\mathcal{A} f = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} P_t f. \tag{159}$$

Theorem 3.17 (Generator of Brownian motion). The Laplacian Δ is the generator of the standard Brownian motion.

Proof of Theorem 3.17. ■

Theorem 3.18 (Semi-group and generator of Ornstein-Uhlenbeck process). Let $(X_t)_{t\geqslant 0}$ be the Ornstein-Uhlenbeck (OU) process:

$$\mathrm{d}X_t = -\frac{X_t}{2}\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t.$$

The semi-group of the OU process is given by:

$$P_t f(x) = \mathbb{E}_{\xi\sim\pi}[f(e^{-t/2}x + \sqrt{1-e^{-t}}\xi)] \qquad t \geqslant 0, \tag{160}$$

where $\pi(x) = (4\pi)^{-n/2}\exp(-|x|^2/4)$ is the stationary measure of (X_t) in the sense that $P_t f \to \pi f$ in $L^2(\pi)$. Moreover, the generator $\mathcal{A}$ of (X_t) is given by:

$$\mathcal{A} f = -\langle \nabla f, \frac{x}{2}\rangle + \Delta f \tag{161}$$

is the drift Laplacian with the potential $\phi(x) = |x|^2/4$.

Remark 3.19. From Theorem 3.18, we see that the backward operator in (155) is the same as the generator, both equal to the drift Laplacian for the OU process. ■

Proof of Theorem 3.18. Obviously $\lim_{t\downarrow 0} P_t f(x) = f(x)$. For $f \in L^2(\pi)$, observe that

$$\begin{aligned} \|P_t f - \pi f\|^2_{L^2(\pi)} &= \mathbb{E}_{X\sim\pi}\left[\mathbb{E}_{\xi\sim\pi}[f(e^{-t/2}X + \sqrt{1-e^{-t}}\xi)]\right]^2 \\ &\leqslant \mathbb{E}_{X\sim\pi}\,\mathbb{E}_{\xi\sim\pi}\left|f(\underbrace{e^{-t/2}}_{\to 0}X + \underbrace{\sqrt{1-e^{-t}}}_{\to 1}\xi)\right|^2 \\ &\to 0 \qquad \text{as } t\to\infty, \end{aligned}$$

where the second inequality is due to Cauchy-Schwarz and the third step is due to the dominated convergence theorem and Fubini's theorem. Thus $P_t f \to \pi f$ in $L^2(\pi)$ in the long-time dynamics.

Now using the chain rule, we compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P_t f(x) &= \mathbb{E}_{\xi\sim\pi}\left[\langle\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi),\, -e^{-t/2}\frac{x}{2}+\frac{e^{-t}}{2\sqrt{1-e^{-t}}}\xi\rangle\right]\\ &= -\frac{e^{-t/2}}{2}\,\mathbb{E}_{\xi\sim\pi}\left[\langle\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi),\, x\rangle\right]\\ &\qquad+\frac{e^{-t}}{2\sqrt{1-e^{-t}}}\,\mathbb{E}_{\xi\sim\pi}\left[\langle\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi),\, \xi\rangle\right].\end{aligned}$$

Note that $\nabla\pi(\xi) = -\frac{\xi}{2}\pi(x)$ and

$$\begin{aligned}\mathrm{div}_\xi\left(\pi(\xi)\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\right) &= \langle\nabla\pi(\xi),\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\rangle\\ &\quad+\pi(\xi)\Delta f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\sqrt{1-e^{-t}}.\end{aligned}$$

So we have

$$\frac{1}{2}\int_{\mathbb{R}^n}\pi(\xi)\langle\xi,\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\rangle\,\mathrm{d}\xi = \int_{\mathbb{R}^n}\pi(\xi)\Delta f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\sqrt{1-e^{-t}}\,\mathrm{d}\xi,$$

i.e.,

$$\frac{1}{2}\,\mathbb{E}_{\xi\sim\pi}\left[\langle\xi,\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\rangle\right] = \sqrt{1-e^{-t}}\,\mathbb{E}_{\xi\sim\pi}\left[\Delta f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\right],$$

which is sometimes referred as *Stein's identity* or the *Gaussian integration-by-parts*. Now, combining all pieces, we see that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P_t f(x) &= -\frac{e^{-t/2}}{2}\,\mathbb{E}_{\xi\sim\pi}\left[\langle\nabla f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi),\, x\rangle\right] + e^{-t}\,\mathbb{E}_{\xi\sim\pi}\left[\Delta f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\right]\\ &= -\frac{1}{2}\,\mathbb{E}_{\xi\sim\pi}\left[\langle\nabla_x f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi),\, x\rangle\right] + \mathbb{E}_{\xi\sim\pi}\left[\Delta_x f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\right]\\ &= -\frac{1}{2}\left\langle\nabla_x\,\mathbb{E}_{\xi\sim\pi}\left[f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\right],\, x\right\rangle + \Delta_x\,\mathbb{E}_{\xi\sim\pi}\left[f(e^{-t/2}x+\sqrt{1-e^{-t}}\xi)\right]\end{aligned}$$

Filling the definition (160) of $P_t f$ into the last expression, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t f(x) = -\frac{1}{2}\langle\nabla P_t f(x),\, x\rangle + \Delta P_t f(x).$$

Since $P_t f\to f$ as $t\downarrow 0$, we have

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} P_t f(x) = -\frac{1}{2}\langle\nabla f(x),\, x\rangle + \Delta f(x),$$

which gives the generator (161) of the OU process. ∎

3.6 Rate of convergence

If $Z = \int_{\mathbb{R}^n} e^{-V(x)} \, \mathrm{d}x < \infty$, then the Fokker-Planck equation has a stationary distribution on $\mathbb{R}^n$:

$$\pi(x) = \frac{1}{Z} e^{-V(x)}, \tag{162}$$

where $0 < Z < \infty$ is a normalization constant. To see this, recall that $F(\rho) = \int \rho \log \rho + \int V \, \mathrm{d}\rho$ and $\frac{\delta F}{\delta \rho}(\pi) = \log \pi + V$. Since $\nabla \pi = Z^{-1} e^{-V}(-\nabla V) = -\pi \nabla V$,

$$\pi \, \nabla \Big(\frac{\delta F}{\delta \rho}(\pi) \Big) = \pi \frac{\nabla \pi}{\pi} + \pi \nabla V = \nabla \pi + \pi \nabla V = 0. \tag{163}$$

Then,

$$\partial_t \pi = \operatorname{div} \Big(\pi \, \nabla \Big(\frac{\delta F}{\delta \rho}(\pi) \Big) \Big) = 0, \tag{164}$$

which implies that π is a stationary point of the Fokker-Planck continuity equation. In this case, the gradient flow of the functional F can be viewed as the gradient flow of the relative entropy between ρ and π (cf. (168) in Definition 3.20 below):

$$F(\rho) + \log Z = \int \rho \Big[\log \Big(\frac{\rho}{e^{-V}} \Big) + \log Z \Big] = \int \rho \log \Big(\frac{\rho}{\pi} \Big) = H(\rho \| \pi) \tag{165}$$

so that

$$\frac{\delta F}{\delta \rho}(\rho) = \frac{\delta (F + \log Z)}{\delta \rho}(\rho) = \frac{\delta H(\rho \| \pi)}{\delta \rho}(\rho). \tag{166}$$

Now suppose the Fokker-Planck equation admits a stationary distribution. Given an initial distribution ρ_0, we would like to ask how fast does the Wasserstein gradient flow $(\rho_t)_{t \geqslant 0}$ converge to the stationary distribution π? In a nutshell, the answer is given by the following statement.

> If the stationary measure π satisfies a logarithmic Sobolev inequality (LSI), then ρ_t converges to π exponentially fast in time t (under numerous distance or divergence measures).

Suppose π is probability measure (i.e., $0 < Z < \infty$). Since the convergence quantities involved depend only on the (first and second) derivatives of the density π, without loss of generality, we may assume $Z = 1$ and

$$\pi = e^{-V}. \tag{167}$$

Definition 3.20 (Relative entropy and relative Fisher information). Let p, q be two probability measures on $\mathbb{R}^n$ such that $q \ll p$. Then the *relative entropy* (or Kullback-Leibler divergence) between p and q is defined as

$$H(p \| q) = \int \log \Big(\frac{\mathrm{d}p}{\mathrm{d}q} \Big) \, \mathrm{d}p. \tag{168}$$

In particular, if $\mathrm{d}x \ll p$ and $\mathrm{d}x \ll q$, then

$$H(p \| q) = \int p \log \Big(\frac{p}{q} \Big) \, \mathrm{d}x. \tag{169}$$

Let π, μ be two probability measures on $\mathbb{R}^n$ such that $\pi \ll \mu$ and $\frac{\mathrm{d}\mu}{\mathrm{d}\pi} = \rho$, then the *relative Fisher information* between μ and π is defined as

$$I(\mu\|\pi) = \int \frac{|\nabla\rho|^2}{\rho}\,\mathrm{d}\pi. \tag{170}$$

We state a powerful information inequality that implies the LSI.

Theorem 3.21 (HWI++ inequality [5]). Let $(\mathbb{R}^n, |\cdot|, \pi)$ be a probability space such that

$$\mathrm{d}\pi(x) = e^{-V(x)}\,\mathrm{d}x, \tag{171}$$

where $V : \mathbb{R}^n \to [0, \infty)$ is a $C^\infty(\mathbb{R}^n)$ function such that $\mathrm{Hess}(V) \succeq \kappa I_n$ for some parameter $\kappa \in \mathbb{R}$. Then for any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ such that $H(\mu_0\|\pi) < \infty$,

$$H(\mu_1\|\pi) - H(\mu_0\|\pi) \leqslant W_2(\mu_0, \mu_1)\sqrt{I(\mu_1\|\pi)} - \frac{\kappa}{2}W_2^2(\mu_0, \mu_1). \tag{172}$$

Theorem 3.21 implies several classical functional and information inequalities.

Corollary 3.22. In the setting of Theorem 3.21 and assume further $\pi \in \mathcal{P}_2(\mathbb{R}^n)$, then we have

1. *HWI inequality (Theorem 3 in [10])*: for any $\nu \in \mathcal{P}_2(\mathbb{R}^n)$,

$$H(\nu\|\pi) \leqslant W_2(\nu, \pi)\sqrt{I(\nu\|\pi)} - \frac{\kappa}{2}W_2^2(\nu, \pi). \tag{173}$$

2. *Talagrand's T_2-inequality*: for any $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ with finite second moment,

$$\frac{\kappa}{2}W_2^2(\nu, \pi) \leqslant H(\nu\|\pi). \tag{174}$$

3. *Logarithmic Sobolev inequality*: if $\kappa > 0$, then for any $\nu \in \mathcal{P}(\mathbb{R}^n)$,

$$H(\nu\|\pi) \leqslant \frac{1}{2\kappa}I(\nu\|\pi). \tag{175}$$

Combining (174) and (175), we have that if $\kappa > 0$, then

$$W_2(\nu, \pi) \leqslant \frac{\sqrt{I(\nu\|\pi)}}{\kappa}, \quad \forall \pi \ll \nu. \tag{176}$$

In Appendix C.1 and C.2, we discuss more details of the LSIs and Talagrand's transportation inequalities.

Proof of Corollary 3.22. Corollary 3.22 follows from Theorem 3.21 with specific choices.

1. Take $\mu_0 = \pi$ and $\mu_1 = \nu$.
2. Take $\mu_0 = \nu$ and $\mu_1 = \pi$.
3. Take $\mu_0 = \pi$ and $\mu_1 = \nu$ to get the HWI inequality in part 1, and then maximize its right-hand side $-\frac{\kappa}{2}x^2 + \sqrt{I(\nu\|\pi)}x$, where the maximizer occurs at $x^* = \frac{1}{\kappa}\sqrt{I(\nu\|\pi)}$. Then a standard approximation argument is sufficient.

$\blacksquare$

There are various ways of measuring the gap between the gradient flow $(\rho_t)_{t\geqslant 0}$ as a solution of the Fokker-Planck equation and the stationary distribution π: total variation (as in Meyn-Tweedie's Markov chain approach), L^2-norm, relative entropy, Wasserstein distance, etc. Below in Theorem 3.23, we state the exponential rate of convergence under the relative entropy.

Theorem 3.23 (Exponential rate of convergence for the Fokker-Planck gradient flow: relative entropy). Let $(\rho_t)_{t\geqslant 0}$ solves the (gradient drift) Fokker-Planck equation:

$$\partial_t \rho_t = \mathrm{div}(\nabla \rho_t + \rho_t \nabla V), \tag{177}$$

where the potential $V > 0$ satisfies $V \in C^2(\mathbb{R}^n)$. If the Bakry-Émery criterion

$$\mathrm{Hess}(V) \succeq \kappa I_n \tag{178}$$

holds for some $\kappa > 0$ (i.e., $\mathrm{Hess}_V(x) \succeq \kappa I_n$ for all $x \in \mathbb{R}^n$), then

$$H(\rho_t \| \pi) \leqslant e^{-2\kappa t} H(\rho_0 \| \pi), \quad t \geqslant 0, \tag{179}$$

where $\pi(x) = e^{-V(x)}$ is the stationary distribution for the Fokker-Planck equation (177).

Note that the Bakry-Émery criterion is a strong convexity (sometimes called the κ-convexity) requirement for the potential V. From the sampling point of view, the Bakry-Émery criterion requires that the stationary distribution π is strictly log-concave to obtain exponential rate of convergence under the relative entropy.

It is known that strictly log-concave probability density satisfies an LSI, and vice versa (cf. Problem 3.17 in [13]). In the celebrated result of Bakry and Émery, a simple sufficient condition is given to ensure that the density satisfies an LSI.

Lemma 3.24 (Bakry-Émery criterion). Let $V : \mathbb{R}^n \to \mathbb{R}$ be a $C^2(\mathbb{R}^n)$ function and $\mathrm{d}\pi = e^{-V}\,\mathrm{d}x$ be a probability measure such that $\mathrm{Hess}(V) \succeq \kappa I_n$ for some $\kappa > 0$. Then π satisfies the LSI with parameter κ:

$$H(\nu \| \pi) \leqslant \frac{1}{2\kappa} I(\nu \| \pi) \tag{180}$$

for all $\nu \in \mathcal{P}(\mathbb{R}^n)$ such that $\pi \ll \nu$.

However, it is an open question whether or not all log-concave densities satisfy an LSI. It is even an open question to ask whether or not all log-concave distributions satisfy a Poincaré inequality with a *dimension-free* constant. This is the Kannan-Lovász-Simonovits (KLS) conjecture [1, 3].

Conjecture 3.25 (Kannan-Lovász-Simonovits conjecture, cf. Conjecture 1.2 in [1]). There exists an absolute constant $C > 0$ such that for any log-concave probability measure ν on $\mathbb{R}^n$, we have

$$\mathrm{Var}_\nu(f) := \mathbb{E}_\nu \, |f - \mathbb{E}_\nu[f]|^2 \leqslant C \lambda_\nu^2 \, \mathbb{E}_\nu \, |\nabla f|^2 \tag{181}$$

for any locally Lipschitz (i.e., Lipschitz on any Euclidean ball) function $f \in L^2(\nu)$, where λ_ν is the square root of the largest eigenvalue of the covariance matrix $\mathbb{E}_{X\sim\nu}[XX^T]$.

Currently, the best known result is that, for an isotropic n-dimensional log-concave distribution, the dimension-dependent constant $C(n)$ is on the order $n^{1/4}$ [9].

We also note that the Bakry-Émery criterion can be replaced by a slightly stronger Otto-Villani criterion (with an additional finite second moment assumption), so that we can obtain both LSI and T_2 inequality.

Lemma 3.26 (Otto-Villani criterion, Theorem 1 in [10]). Let $V : \mathbb{R}^n \to \mathbb{R}$ be a $C^2(\mathbb{R}^n)$ function and $\mathrm{d}\pi = e^{-V}\,\mathrm{d}x$ be a probability measure with a finite second moment such that $\mathrm{Hess}(V) \succeq \kappa I_n$ for some $\kappa > 0$. Then π satisfies the LSI and Talagrand's T_2 inequality, both with parameter κ:

$$H(\nu\|\pi) \leqslant \frac{1}{2\kappa} I(\nu\|\pi) \quad \text{and} \quad \frac{\kappa}{2} W_2^2(\nu,\pi) \leqslant H(\nu\|\pi) \tag{182}$$

for all $\nu \in \mathcal{P}(\mathbb{R}^n)$ such that $\pi \ll \nu$.

With the (slightly) stronger Otto-Villani criterion, exponential rate of convergence also holds under the Wasserstein distance [2]. In fact, rate of convergence for (more general) non-gradient drift Fokker-Planck equation is established in [2].

Theorem 3.27 (Exponential rate of convergence for the Fokker-Planck gradient flow: Wasserstein distance). Let $(\rho_t)_{t\geqslant 0}$ solves the (gradient drift) Fokker-Planck equation (177) with the stationary distribution $\mathrm{d}\pi = e^{-V}\,\mathrm{d}x$ satisfying the Otto-Villani criterion for some $\kappa > 0$ (i.e., $\mathrm{d}\pi = e^{-V}\,\mathrm{d}x$ has finite second moment such that $V \in C^2(\mathbb{R}^n)$ and $\mathrm{Hess}(V) \succeq \kappa I_n$). Then,

$$W_2(\rho_t, \pi) \leqslant e^{-\kappa t} W_2(\rho_0, \pi). \tag{183}$$

Proof of Theorem 3.27. Using the relation between the PDE and SDE, solution to the (gradient drift) Fokker-Planck equation (177) is the density of the stochastic process $(X_t)_{t\geqslant 0}$ solving the following SDE:

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{184}$$

where $(W_t)_{t\geqslant 0}$ is the standard Brownian motion on $\mathbb{R}^n$ and the initial datum has distribution ρ_0 (cf. Section **??** for more details of the equivalence).

Let μ_0 and ν_0 be two probability measures on $\mathbb{R}^n$, and (X_0, Y_0) be a coupling at the initial time with the marginal distributions $X_0 \sim \mu_0$ and $Y_0 \sim \nu_0$ such that

$$\mathbb{E}\,|X_0 - Y_0|^2 = W_2^2(\mu_0, \nu_0). \tag{185}$$

Then we run two coupled copies of the SDE with $(X_t)_{t\geqslant 0}$ (resp. $(Y_t)_{t\geqslant 0}$) as the solution to (184) starting from $X_0 \sim \mu_0$ (resp. $Y_0 \sim \nu_0$), both driven by the same Brownian motion $(W_t)_{t\geqslant 0}$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\,|X_t - Y_t|^2 = -2\,\mathbb{E}\,\big\langle X_t - Y_t,\, \nabla V(X_t) - \nabla V(Y_t)\big\rangle. \tag{186}$$

Note that for any $x, y \in \mathbb{R}^n$,

$$V(x) = V(y) + \langle x - y, \nabla V(y)\rangle + \frac{1}{2}(x-y)^T \mathrm{Hess}_V(z)(x-y), \tag{187}$$

$$V(y) = V(x) + \langle y - x, \nabla V(x)\rangle + \frac{1}{2}(y-x)^T \mathrm{Hess}_V(z')(y-x), \tag{188}$$

where z, z' are on the line segment joining x and y. Adding the last two equalities, we get

$$\langle x-y, \nabla V(x) - \nabla V(y)\rangle = \frac{1}{2}(x-y)^T[\mathrm{Hess}_V(z) + \mathrm{Hess}_V(z')](x-y). \tag{189}$$

If $\mathrm{Hess}(V) \succeq \kappa I_n$ for some $\kappa > 0$, then

$$\big\langle x-y,\, \nabla V(x) - \nabla V(y)\big\rangle \geqslant \kappa |x-y|^2, \quad \forall x, y \in \mathbb{R}^n. \tag{190}$$

Then Grönwall's lemma yields that

$$\mathbb{E}\,|X_t - Y_t|^2 \leqslant e^{-2\kappa t}\,\mathbb{E}\,|X_0 - Y_0|^2. \tag{191}$$

By definition of the Wasserstein distance,

$$W_2^2(\mu_t, \nu_t) \leqslant \mathbb{E}\,|X_t - Y_t|^2. \tag{192}$$

Now we have

$$W_2^2(\mu_t, \nu_t) \leqslant e^{-2\kappa t} W_2^2(\mu_0, \nu_0), \tag{193}$$

which gives an exponential contraction between *any* two solutions μ_t and ν_t to the Fokker-Planck equation (177). In particular, (183) follows from choosing ν_0 (and thus all $\nu_t, t \geqslant 0$) as the stationary solution $\pi = e^{-V}$. ∎

For the Mehler flow, recall that the stationary distribution is given by

$$\pi(x) = (4\pi)^{-n/2} \exp\Big(-\frac{|x|^2}{4}\Big), \tag{194}$$

which is a $\sqrt{2}$-rescaled standard Gaussian distribution γ on $\mathbb{R}^n$. By the Gaussian LSI for γ, we know that π satisfies a similar LSI and thus the Mehler flow has the exponential rate of convergence to its stationary distribution π.

Corollary 3.28 (Exponential rate of convergence for the Mehler flow). We have

$$H(\rho_t\|\pi) \leqslant e^{-t} H(\rho_0\|\pi) \quad \text{and} \quad W_2(\rho_t, \pi) \leqslant e^{-t/2} W_2(\rho_0, \pi), \quad t \geqslant 0, \tag{195}$$

where π is the stationary distribution the Mehler flow $(\rho_t)_{t\geqslant 0}$ with $V(x) = \frac{|x|^2}{4}$. In particular, ρ_t converges weakly to π (in distribution) as $t \to \infty$.

Proof of Corollary 3.28. For the Mehler flow, $V(x) = \frac{|x|^2}{4}, \nabla V = \frac{x}{2}$, and $\mathrm{Hess}(V) = \frac{1}{2} I_n$. So $\kappa = \frac{1}{2}$. Then Corollary 3.28 follows from Theorem 3.23 and Theorem 3.27. ∎

Corollary 3.28 is a quantitative version of Boltzmann's H-theorem stating that the total entropy of an isolated system can never decrease over time (i.e., the second law of thermodynamics):

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\rho_t\|\pi) = -I(\rho_t\|\pi) \leqslant 0 \quad \text{with } \pi = \frac{e^{-V}}{Z}. \tag{196}$$

4 Optimal transport

4.1 Constant-speed geodesics in W_p

Definition 4.1 (Length of curve). The *length* of a curve $\omega : [0, 1] \to X$ in a metric space (X, d) is defined as

$$\mathrm{Length}(\omega) = \sup \left\{ \sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1})) : n \geqslant 1,\, 0 = t_0 < t_1 < \cdots < t_n = 1 \right\}. \tag{197}$$

Remark 4.2 (Absolutely continuous curve has finite length). By Definition 3.3, if $\omega \in \mathrm{AC}(X)$, then there exists an $g \in L^1([0, 1])$ (i.e., g has the bounded variation) such that

$$d(\omega(t_k), \omega(t_{k+1})) \leqslant \int_{t_k}^{t_{k+1}} g(\tau)\, \mathrm{d}\tau.$$

So we have

$$\sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1})) \leqslant \int_0^1 g(\tau)\, \mathrm{d}\tau \implies \mathrm{Length}(\omega) \leqslant \int_0^1 g(\tau)\, \mathrm{d}\tau < \infty.$$

■

Lemma 4.3 (Derivative of length is metric derivative). If $\omega \in \mathrm{AC}(X)$ is an absolutely continuous curve in a metric space (X, d), then

$$\mathrm{Length}(\omega) = \int_0^1 |\omega'|(t)\, \mathrm{d}t, \tag{198}$$

where $|\omega'|(t)$ is the metric derivative of the curve ω

$$|\omega'|(t) = \lim_{h \to 0} \frac{d(\omega(t+h), \omega(t))}{|h|}. \tag{199}$$

Proof of Lemma 4.3. Since $\omega \in \mathrm{AC}(X)$, after the reparameterization in Remark 3.5, we can assume that ω is 1-Lipschitz continuous. So by the Rademacher theorem (cf. Theorem 3.6), the metric derivative $|\omega'|(t)$ exists and we have $\mathrm{Length}(\omega) \leqslant \int_0^1 |\omega'|(t)\, \mathrm{d}t$. ■

Definition 4.4 (Length space and geodesic space). Let $\omega : [0, 1] \to X$ be an absolutely continuous curve in a metric space (X, d). The curve ω is said to be a *geodesic* between $x_0, x_1 \in X$ if it minimizes the length among all absolutely continuous curves such that $\omega(0) = x_0$ and $\omega(1) = x_1$.

A space (X, d) is said to be a *length space* if

$$d(x, y) = \inf \{\mathrm{Length}(\omega) : \omega(0) = x_0, \omega(1) = x_1, \omega \in \mathrm{AC}(X)\}. \tag{200}$$

A space (X, d) is said to be a *geodesic space* if

$$d(x, y) = \min \{\mathrm{Length}(\omega) : \omega(0) = x_0, \omega(1) = x_1, \omega \in \mathrm{AC}(X)\}. \tag{201}$$

Definition 4.5 (Constant-speed geodesic). Let (X, d) be a length space. A curve $\omega : [0, 1] \to X$ is a *constant-speed geodesic* between $\omega(0)$ and $\omega(1)$ in X if

$$d(\omega(t), \omega(s)) = |t - s| \, d(\omega(0), \omega(1)) \qquad \forall \, t, s \in [0, 1]. \tag{202}$$

Note that a constant-speed geodesic ω is indeed a geodesic. Take any $0 = t_0 < t_1 < \cdot < t_n = 1$ and any other curve $\tilde{\omega} : [0, 1] \to X$ such that $\tilde{\omega}(0) = \omega(0)$ and $\tilde{\omega}(1) = \omega(1)$. Then we have

$$\sum_{k=0}^{n-1} d(\tilde{\omega}(t_k), \tilde{\omega}(t_{k+1})) \geqslant d(\tilde{\omega}(t_0), \tilde{\omega}(t_n)) = d(\tilde{\omega}(0), \tilde{\omega}(1)) = d(\omega(0), \omega(1)),$$
$$\sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1})) = \sum_{k=0}^{n-1} (t_{k+1} - t_k) d(\omega(0), \omega(t_1)) = d(\omega(0), \omega(t_1)),$$

which imply that

$$\mathrm{Length}(\tilde{\omega}) \geqslant d(\omega(0), \omega(1)) = \mathrm{Length}(\omega).$$

Lemma 4.6 (Characterization of constant-speed geodesic in metric space). The $p > 1$ and $\omega : [0, 1] \to X$ connecting x_0 and x_1 in X. Then the followings are equivalent.

1. ω is a constant-speed geodesic.
2. $\omega \in \mathrm{AC}(X)$ such that $|\omega'|(t) = d(\omega(0), \omega(1))$ for almost everywhere $t \in [0, 1]$.
3. ω solves $\min \left\{ \int_0^1 |\tilde{\omega}'|^p(t) \, \mathrm{d}t : \tilde{\omega}(0) = x_0, \tilde{\omega}(1) = x_1 \right\}$.

Theorem 4.7 (Constant-speed geodesic in W_p). Let $p \geqslant 1$, $\Omega \subset \mathbb{R}^n$ be a convex subset, and $\mu, \nu \in \mathcal{P}_p(\Omega)$. Suppose $\gamma \in \Gamma(\mu, \nu)$ is an optimal transport plan for the cost $|x - y|^p$. Define $\pi_t : \Omega \times \Omega \to \Omega$ as

$$\pi_t(x, y) = (1 - t)x + ty. \tag{203}$$

Then the curve $\mu_t = (\pi_t)_\sharp \gamma$ is a constant-speed geodesic in $(\mathcal{P}_p(\Omega), W_p)$ connecting $\mu_0 = \mu$ and $\mu_1 = \nu$.

Remark 4.8 (McCann's interpolation). If μ has a density w.r.t. the Lebesgue measure $\mathrm{d}x$ on $\mathbb{R}^n$, then, by Theorem 4.7, there is an optimal transport map T from μ to ν such that

$$\mu_t = ((1 - t)\mathrm{id} + tT)_\sharp \, \mu \tag{204}$$

is a constant-speed geodesic in W_p. This implies that the p-Wasserstein space $(\mathcal{P}_p(\Omega), W_p)$ is a geodesic space. The curve (204) is called *McCann's interpolation*. In particular, $\mu_{1/2}$ is called the *barycenter* of μ and ν.

Proof of Theorem 4.7. It suffices to show that $W_p(\mu_t, \mu_s) \leqslant (t - s)W_p(\mu, \nu)$ for all $0 \leqslant s < t \leqslant 1$ because this inequality and the triangle inequality

$$W_p(\mu, \nu) \leqslant W_p(\mu_0, \mu_s) + W_p(\mu_s, \mu_t) + W_p(\mu_t, \mu_1) \leqslant [(s - 0) + (t - s) + (1 - t)]W_p(\mu, \nu) = W_p(\mu, \nu)$$

imply that all inequalities are equalities. Thus $W_p(\mu_s, \mu_t) = (t - s)W_p(\mu, \nu)$ for all $0 \leqslant s < t \leqslant 1$. Recall $\mu_t = (\pi_t)_\sharp \gamma$, where $\pi_t(x, y) = (1 - t)x + ty$. Let

$$\gamma_s^t := (\pi_s, \pi_t)_\sharp \gamma = ((\pi_s)_\sharp \gamma, (\pi_t)_\sharp \gamma) \in \Gamma(\mu_s, \mu_t) \tag{205}$$

be a coupling of μ_s and μ_t. Then,

$$\begin{aligned}
W_p(\mu_s,\mu_t) &\leqslant \left(\int_\Omega |x-y|^p \,\mathrm{d}\gamma_s^t\right)^{1/p} = \left(\int_\Omega |x-y|^p \,\mathrm{d}(\pi_s,\pi_t)_\sharp\gamma\right)^{1/p} \\
&=_{(*)} \left(\int_\Omega |\pi_s(x,y)-\pi_t(x,y)|^p \,\mathrm{d}\gamma(x,y)\right)^{1/p} \\
&= \left(\int_\Omega |(1-s)x+sy-(1-t)x-ty|^p \,\mathrm{d}\gamma(x,y)\right)^{1/p} \\
&= \left(\int_\Omega |t-s|^p|x-y|^p \,\mathrm{d}\gamma(x,y)\right)^{1/p} \\
&= |t-s| \left(\int_\Omega |x-y|^p \,\mathrm{d}\gamma(x,y)\right)^{1/p} = (t-s)W_p(\mu,\nu),
\end{aligned}$$

where $(*)$ follows from change of variables. ■

4.2 Benamou-Brenier formulation

Now we are ready to formulate a dynamic version of the Kantorovich optimal transport problem as an optimal flow problem. This is referred as the *Benamou-Brenier formulation.*

Recall that $(\mathcal{P}_p(\Omega), W_p)$ is a geodesic space. For $\mu,\nu \in \mathcal{P}_p(\Omega)$, we can take a constant-speed geodesic $(\mu_t)_{t\in[0,1]}$ connecting μ_0 and $\nu=\mu_1$. By Lemma 4.6, $(\mu_t)_{t\in[0,1]}$ is an absolutely continuous curve such that for almost everywhere $t\in[0,1]$,

$$|\mu'|(t) = W_p(\mu_0,\mu_1) = W_p(\mu,\nu)$$

and μ_t solves

$$\min\left\{\int_0^1 |\tilde{\mu}'|^p(t)\,\mathrm{d}t : \tilde{\mu}(0)=x_0, \tilde{\mu}(1)=x_1, \tilde{\mu}\in \mathrm{AC}(\mathcal{P}_p(\Omega))\right\}.$$

Theorem 4.9 (Benamou-Brenier formula). We have

$$W_p^p(\mu,\nu) = \min\left\{\int_0^1 \|\mathbf{v}_t\|_{L^p(\mu_t)}^p\,\mathrm{d}t : \tilde{\mu}(0)=\mu, \tilde{\mu}(1)=\nu, \partial_t\tilde{\mu}_t + \mathrm{div}(\tilde{\mu}_t\mathbf{v}_t)=0\right\}. \tag{206}$$

Proof of Theorem 4.9. ■

4.3 Caffarelli contration theorem

Definition 4.10 (Brenier map). Let μ and ν be two probability measures on $\mathbb{R}^n$. The *Brenier map* (i.e., the optimal transport map) $y : (\mathbb{R}^n,\mu) \to (\mathbb{R}^n,\nu)$ satisfies the following two properties:

1. *cost-minimizing:* y minimizes the cost functional $\int_{\mathbb{R}^n} c(x,y(x))\,\mathrm{d}\mu(x)$;
2. *measure-preserving:* for any test function $\psi:\mathbb{R}^n\to\mathbb{R}$, one has
$$\int_{\mathbb{R}^n}\psi(y)\,\mathrm{d}\nu(y) = \int_{\mathbb{R}^n}\psi(y(x))\,\mathrm{d}\mu(x), \tag{207}$$
i.e., ν is the pushforward measure of μ by y, denote as $\nu = y_\sharp\mu$.

Classical optimal transport theory imply that the Brenier map exists and unique if $\mu \gg \mathrm{d}x$. In addition, if μ and ν has smooth densities f and g w.r.t. the Lebesgue measure $\mathrm{d}x$ on $\mathbb{R}^n$, respectively, such that $f, g > 0$, then the Brenier map y is given by a gradient of a convex function, i.e.,

$$y(x) = \nabla\varphi(x) \tag{208}$$

for some convex and smooth potential $\varphi : \mathbb{R}^n \to \mathbb{R}$. Since y is measure-preserving, by change of variables, we have

$$\int_{\mathbb{R}^n} \psi(\nabla\varphi(x)) f(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \psi(y) g(y)\,\mathrm{d}y = \int_{\mathbb{R}^n} \psi(\nabla\varphi(x)) g(\nabla\varphi(x)) \det(\nabla^2\varphi(x))\,\mathrm{d}x.$$

This gives the *Monge-Ampère equation* (in the distribution sense):

$$\det(\nabla^2\varphi(x)) = \frac{f(x)}{g(\nabla\varphi(x))}. \tag{209}$$

Remark 4.11 (Regularity of φ). If $f, g \in \mathcal{C}^2(\Omega)$ where $\Omega \subset \mathbb{R}^n$ is a convex subset, then $\varphi \in \mathcal{C}^{2,\alpha}(\Omega)$, where φ satisfies the Monge-Ampère equation (209) in the classical sense. This implies that φ is a strongly convex (i.e., $\nabla^2\varphi(x) > 0$) *strong/classical*) solution of the Monge-Ampère equation. ■

A theorem of Caffarelli in [4] shows that the Brenier map pushing forward a source Gaussian measure to a more log-concave target measure contracts Euclidean distance.

Theorem 4.12 (Caffarelli contration theorem). Let μ and ν be two probability measures on $\mathbb{R}^n$ with Lebesgue densities f and g, respectively, i.e., $\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$ and $\mathrm{d}\nu(x) = g(x)\,\mathrm{d}x$. Suppose $f(x) = e^{-Q(x)}$, where $Q(x) = \frac{1}{2}x^T Ax$ for some positive semi-definite matrix $A = (a_{ij})_{i,j=1}^n$ (not depend on x). Let $g(x) = e^{-Q(x)-F(x)}$, where $F : \mathbb{R}^n \to \mathbb{R}$ is convex. Let $x \mapsto y(x)$ be the Brenier map from μ to ν. Then we have

$$|y(x_1) - y(x_2)| \leqslant |x_1 - x_2| \tag{210}$$

for all $x_1, x_2 \in \mathbb{R}^n$, i.e., y is a contraction (or 1-Lipschitz function).

To prove Caffarelli contraction in Theorem 4.12, we first give a formal proof and then a rigorous proof.

Formal proof of Theorem 4.12. Let $\rho(B) = \operatorname{logdet}(B)$ for an $n \times n$ positive definite matrix B. By the Monge-Ampère equation (209),

$$\rho(\nabla^2\varphi(x)) = \operatorname{logdet}(\nabla^2\varphi(x)) = \log f(x) - \log g(\nabla\varphi(x)). \tag{211}$$

Assume $\varphi \in \mathcal{C}^4$ such that $\nabla^2\varphi > 0$ is positive definite (as a matrix inequality) and $y = \nabla\varphi$. Denote

$$\begin{aligned}
\varphi_i &:= \partial_i\varphi = \frac{\partial}{\partial x_i}\varphi, \quad \nabla_i\rho(\nabla^2\varphi) := \partial_i\rho(\nabla^2\phi) = \frac{\partial}{\partial x_i}\rho(\nabla^2\phi), \\
\varphi_{ij} &:= \partial^2_{ij}\varphi = \frac{\partial^2}{\partial x_i x_j}\varphi, \quad \nabla^2_{ij}\rho(\nabla^2\varphi) := \partial^2_{ij}\rho(\nabla^2\phi) = \frac{\partial^2}{\partial x_i x_j}\rho(\nabla^2\phi).
\end{aligned}$$

Using the chain rule on the matrix $B = (b_{ij})_{i,j=1}^n$, we have

$$\begin{aligned}
\nabla_1 \rho(\nabla^2 \varphi) &= \sum_{i,j=1}^n \partial_{ij} \rho(\nabla^2 \varphi) \nabla^2_{ij} \varphi_1, \\
\nabla^2_{11} \rho(\nabla^2 \varphi) &= \sum_{i,j=1}^n \partial_{ij} \rho(\nabla^2 \varphi) \nabla^2_{ij} \varphi_{11} + \sum_{i,j=1}^n \sum_{k,l=1}^n \partial^2_{ij,kl} \rho(\nabla^2 \varphi) \nabla^2_{ij} \varphi_1 \nabla^2_{kl} \varphi_1.
\end{aligned} \tag{212}$$

Note that $\partial_{ij} \rho(B) = \frac{\partial}{\partial B} \rho(B) = B^{-1}$, i.e., $\frac{\partial}{\partial b_{ij}} \rho = b^{ji}$ where $B^{-1} = (b^{ji})_{i,j=1}^n$ if $\det(B) \neq 0$. Assume that φ_{11} achieves a maximum in $\mathbb{R}^n$. At the maximum point $x_0 \in \mathbb{R}^n$ of φ, we have $\nabla^2_{ij} \varphi_{11}(x_0) \leqslant 0$ because $\varphi \in \mathcal{C}^4$. Applying Lemma 4.13 to (212), we have

$$\begin{aligned}
\nabla^2_{11} \rho(\nabla^2 \varphi(x_0)) &= \underbrace{\sum_{i,j=1}^n \partial_{ij} \rho(\nabla^2 \varphi(x_0))}_{=(\nabla \varphi(x_0))^{-1} > 0} \underbrace{\nabla^2_{ij} \varphi_{11}(x_0)}_{\leqslant 0} + \sum_{i,j=1}^n \sum_{k,l=1}^n \underbrace{\partial^2_{ij,kl} \rho(\nabla^2 \varphi(x_0))}_{<0} \nabla^2_{ij} \varphi_1(x_0) \nabla^2_{kl} \varphi_1(x_0) \\
&\leqslant 0.
\end{aligned}$$

Then this inequality and the Monge-Ampère equation (211) imply that

$$\begin{aligned}
\nabla^2_{11} \rho(\nabla^2 \varphi(x_0)) &= \nabla^2_{11} \log f(x_0) - \nabla^2_{11} \log g(\nabla \varphi(x_0)) \\
&= -\nabla^2_{11} Q(x_0) + \nabla^2_{11} Q(\nabla \varphi(x_0)) + \nabla^2_{11} F(\nabla \varphi(x_0)) \\
&\leqslant 0.
\end{aligned}$$

Since φ_{11} achieves the maximum at x_0 in all directions $\varphi_{\alpha\alpha}, \alpha = 1, \ldots, n$, we can rotate in a suitable basis such that we can diagonalize the Hessian matrix

$$\nabla^2 \varphi(x_0) = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{pmatrix}$$

with $\lambda_1 = \varphi_{11}(x_0)$, $\varphi_{11i}(x_0) = 0$ for all $i = 1, \ldots, n$, and $\varphi_{i1}(x_0) = 0$ for all $i \neq 1$. Note that

$$\begin{aligned}
\nabla_1 F(\nabla \varphi(x)) &= \sum_{i=1}^n \partial_i F(\nabla \varphi(x)) \, \varphi_{i1}(x), \\
\nabla^2_{11} F(\nabla \varphi(x)) &= \sum_{i,j=1}^n \partial^2_{ij} F(\nabla \varphi(x)) \, \varphi_{i1}(x) \, \varphi_{j1}(x) + \sum_{i=1}^n \partial_i F(\nabla \varphi(x)) \, \varphi_{i11}(x) \geqslant \sum_{i=1}^n \partial_i F(\nabla \varphi(x)) \, \varphi_{i11}(x),
\end{aligned}$$

where the last inequality is due to the convexity of F. So we have

$$\nabla^2_{11} F(\nabla \varphi(x_0)) \geqslant \sum_{i=1}^n \partial_i F(\nabla \varphi(x_0)) \, \varphi_{i11}(x_0) = 0.$$

Thus, we get

$$\nabla^2_{11} Q(x_0) \geqslant \nabla^2_{11} Q(\nabla \varphi(x_0)) + \nabla^2_{11} F(\nabla \varphi(x_0)) \geqslant \nabla^2_{11} Q(\nabla \varphi(x_0)). \tag{213}$$

Recall that $Q(x) = \frac{1}{2}x^T Ax$ and $\nabla^2_{11}Q(x) = a_{11}$ where $A = (a_{ij})_{i,j=1}^n$. Then,

$$\nabla_1 Q(\nabla\varphi(x)) = \sum_{i,j=1}^n a_{ij}\,(\nabla\varphi(x))_j\,\varphi_{i1}(x) = \sum_{i,j=1}^n a_{ij}\,\varphi_j(x)\,\varphi_{i1}(x),$$
$$\nabla^2_{11} Q(\nabla\varphi(x)) = \sum_{i,j=1}^n a_{ij}\left[\varphi_{j1}(x)\,\varphi_{i1}(x) + \varphi_j(x)\,\varphi_{i11}(x)\right].$$

Then we have

$$\begin{aligned}\nabla^2_{11} Q(\nabla\varphi(x_0)) &= \sum_{i,j=1}^n a_{ij}\left[\varphi_{j1}(x_0)\,\varphi_{i1}(x_0) + \varphi_j(x_0)\,\varphi_{i11}(x_0)\right] \\ &= \sum_{i,j=1}^n a_{ij}\varphi_{j1}(x_0)\,\varphi_{i1}(x_0) = a_{11}\varphi^2_{11}(x_0).\end{aligned}$$

Now filling all pieces into (213), we get $a_{11}\varphi^2_{11}(x_0) \leqslant a_{11}$. Since $a_{11} > 0$, we conclude that $\phi^2_{11}(x_0) \leqslant 1$, which implies that the eigenvalues of the Hessian matrix $\nabla^2\varphi(x)$ are uniformly bounded in $[0, 1]$ on $\mathbb{R}^n$. This means that the Brenier map $y = \nabla\varphi$ is a 1-Lipschitz function. ■

Lemma 4.13 (Concavity of logdet function). The function $\rho(B) = \operatorname{logdet}(B)$ for $n \times n$ positive definite matrix $B > 0$ is strictly concave on positive symmetric matrices, i.e.,

$$\left(\frac{\partial^2\rho}{\partial b_{ij} b_{kl}}\right)_{i,j,k,l=1}^n < 0 \tag{214}$$

is a negative definite symmetric matrix (as a matrix inequality).

Proof of Lemma 4.13. We first prove the 2×2 case. After diagonalization, we may assume that

$$B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad \text{with } \lambda_1, \lambda_2 > 0.$$

For $x, y, z \in \mathbb{R}$, define

$$\varrho(x, y, z) = \operatorname{logdet}\begin{pmatrix} \lambda_1 + x & z \\ z & \lambda_2 + y \end{pmatrix} = \log\left((\lambda_1 + x)(\lambda_2 + y) - z^2\right).$$

Then elementary calculation yields that

$$\left.\frac{\partial^2\rho}{\partial x^2}\right|_{(0,0,0)} = -\frac{1}{\lambda_1^2}, \quad \left.\frac{\partial^2\rho}{\partial y^2}\right|_{(0,0,0)} = -\frac{1}{\lambda_2^2}, \quad \left.\frac{\partial^2\rho}{\partial z^2}\right|_{(0,0,0)} = -\frac{2}{\lambda_1\lambda_2},$$

and all mixed second-order partial derivatives are zeros. For the general $n \times n$ case, after diagonalization, we similarly can assume B is a diagonal matrix with positive eigenvalues $\lambda_1, \dots, \lambda_n$. Define

$$\varrho((x_{ij})_{1\leqslant i<j\leqslant n}) := \operatorname{logdet}\Big(B + \sum_{i=1}^n x_{ii}E_{ii} + \sum_{1\leqslant i<j\leqslant n} x_{ij}(E_{ij} + E_{ji})\Big),$$

where $E_{ij} = e_i e_j^T$ and $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ with the only i-th position being 1 and all other entries being 0. Then,

$$\frac{\partial^2 \rho}{\partial x_{ii}^2}\bigg|_{(0,\dots,0)} = -\frac{1}{\lambda_i^2}, \qquad \frac{\partial^2 \rho}{\partial x_{ij}^2}\bigg|_{(0,\dots,0)} = -\frac{2}{\lambda_i \lambda_j} \text{ for } i \neq j,$$

and all mixed second-order partial derivatives are zeros. ■

4.3.1 Rigorous proof

There are several places in the formal proof of Theorem 4.12 to be made rigorous. This involves some extra geometric argument. The idea is to replace $\nabla^2\varphi$ by the *second difference* of φ.

Rigorous proof of Theorem 4.12. Define the second difference of φ as

$$\begin{aligned}\delta\varphi(x) := \delta^2\varphi(x) &= [\varphi(x+he) - \varphi(x)] - [\varphi(x) - \varphi(x-he)] \\ &= \varphi(x+he) + \varphi(x-he) - 2\varphi(x),\end{aligned}$$

where e is a unit vector in $\mathbb{R}^n$ and $h > 0$. Since

$$\lim_{h\downarrow 0} \frac{\delta\varphi(x)}{h^2} = \langle \nabla^2\varphi(x) e, e\rangle,$$

it suffices to show that

$$\delta\varphi(x) \leqslant h^2. \tag{215}$$

<u>Step 1:</u> *show that for each fixed h and e, $\delta\varphi(x) \to 0$ as $|x| \to \infty$.*

Fix an $h > 0$ and unit vector $e \in \mathbb{R}^n$ (i.e., $|e| = 1$). By approximation, we can modify F such that $F = +\infty$ on $\mathbb{R}^n \setminus B_R$ for some large enough R, i.e.,

$$g(y) := g_R(x) = e^{-Q(y)-F(y)} \begin{cases} > 0 & \text{on } y \in B_R \\ \to 0 & \text{as } y \to \partial B_R \end{cases} \quad \text{and} \quad y_R = \nabla\varphi, \tag{216}$$

where φ is the Brenier map from $f\,\mathrm{d}x$ to $g_R\,\mathrm{d}x$. In the sequel, we shall denote $g(x) := g_R(x)$ $y(x) := y_R(x)$.

Lemma 4.14. In the setting of Theorem 4.12 and let $y_R(x)$ be defined in (216). Then $y_R(x) \to \frac{x}{|x|}R$ uniformly as $|x| \to \infty$.

Let $x_t = x_0 + te$. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}(\varphi(x_t) - \varphi(x_0)) = \langle \nabla\varphi(x_t), e\rangle,$$

fundamental theorem of calculus yields that

$$\begin{aligned}\delta\varphi(x_0) &= [\varphi(x_0+he) - \varphi(x_0)] - [\varphi(x_0) - \varphi(x_0-he)] \\ &= \int_0^h \langle \nabla\varphi(x_t), e\rangle\,\mathrm{d}t - \int_0^h \langle \nabla\varphi(x_{-t}), e\rangle\,\mathrm{d}t \\ &= \int_0^h \langle \nabla\varphi(x_0+te) - \nabla\varphi(x_0-te), e\rangle\,\mathrm{d}t. \end{aligned} \tag{217}$$

By Lemma 4.14, we have

$$\nabla\varphi(x_0 \pm te) = y(x_0 \pm te) \approx \frac{x_0 \pm te}{|x_0 \pm te|} R \quad \text{uniformly in } 0 \leqslant t \leqslant h,$$

as $|x_0 \pm te| \to \infty$, which implies that

$$\nabla\varphi(x_0 \pm te) \approx \frac{x_0}{|x_0|} R \quad \text{as } |x_0| \to \infty.$$

Then,

$$\langle \nabla\varphi(x_0 + te) - \nabla\varphi(x_0 - te), e\rangle \approx \left\langle \frac{2te}{|x_0|} R, e \right\rangle = \frac{2tR}{|x_0|} \to 0$$

uniformly in $0 \leqslant t \leqslant h$ and $|e| = 1$, as $|x_0| \to \infty$. Integrating this to get

$$\delta\varphi(x_0) \to 0 \quad \text{as } |x_0| \to \infty.$$

Thus $\delta\varphi$ achieves its maximum over all (x_0, e) such that $|e| = 1$.
Step 2: *mimic the formal proof by replacing $\nabla^2\varphi$ with $\delta\varphi$ and bound $\delta\varphi$.*

Now, we fix a point $x_0 \in \mathbb{R}^n$ and a direction e, at which $\delta\varphi$ achieves its maximum. Define

$$\begin{aligned} \theta(x) &:= \log f(x) - \log g(\nabla\varphi(x)), \\ \rho(\nabla^2\varphi(x)) &:= \operatorname{logdet}(\nabla^2\varphi(x)). \end{aligned}$$

By the Monge-Ampère equation (209), we have $\theta(x) = \rho(\nabla^2\varphi(x))$. Since ρ is concave (cf. Lemma 4.13), $\rho(B) - \rho(A) \leqslant \langle \nabla\rho(A), B - A\rangle$ for any $A, B > 0$. Then,

$$\begin{aligned} \rho(\nabla^2\varphi(x_0 \pm he)) - \rho(\nabla^2\varphi(x_0)) &\leqslant \langle \nabla\rho(\nabla^2\varphi(x_0)), \nabla^2\varphi(x_0 \pm he) - \nabla^2\varphi(x_0)\rangle \\ &= \sum_{i,j=1}^n M_{ij}(\nabla^2\varphi(x_0)) \left(\varphi_{ij}(x_0 \pm he) - \varphi_{ij}(x_0)\right), \end{aligned}$$

where $[M_{ij}(\nabla^2\varphi(x_0))]_{i,j=1}^n = [\nabla^2\varphi(x_0)]^{-1} > 0$ as a matrix inequality. Thus, we get

$$\theta(x_0 \pm he) - \theta(x_0) \leqslant \sum_{i,j=1}^n M_{ij}(\nabla^2\varphi(x_0)) \left(\varphi_{ij}(x_0 \pm he) - \varphi_{ij}(x_0)\right),$$

which implies that

$$\begin{aligned} \delta\theta(x_0) &= [\theta(x_0 + he) - \theta(x_0)] + [\theta(x_0 - he) - \theta(x_0)] \\ &\leqslant \sum_{i,j=1}^n M_{ij}(\nabla^2\varphi(x_0))\delta\varphi_{ij}(x_0) \\ &= \langle \underbrace{M(\nabla^2\varphi(x_0))}_{>0}, \underbrace{\delta\varphi(x_0)}_{\leqslant 0}\rangle \leqslant 0, \end{aligned} \tag{218}$$

because $\nabla^2\delta\varphi(x_0) \leqslant 0$ as x_0 is a maximizer of $\delta\varphi$ (i.e., the second derivative test for $\delta\varphi$).

On the other hand, recall

$$\theta(x) = \log f(x) - \log g(\nabla\varphi(x)) = -Q(x) + Q(y(x)) + F(y(x)), \tag{219}$$

where $y(x) = \nabla\varphi(x)$. Since $Q(x) = B(x,x)$ is a symmetric bilinear form, we have for any $z \in \mathbb{R}^n$,

$$\begin{aligned}
\delta Q(z) &= [Q(z+he) - Q(z)] + [Q(z-he) - Q(z)] \\
&= B(z+he, z+he) + B(z-he, z-he) - 2B(z,z) \\
&= 2h^2 B(e,e),
\end{aligned}$$

which does not depend on z. So we can write

$$\delta Q(x) = 2h^2 B(e,e) \quad \text{with } x \mapsto x \pm he.$$

Because $\delta\varphi(x)$ achieves maximum at x_0, by the first derivative test, we have

$$\nabla\delta\varphi(x_0) = \nabla\varphi(x_0 + eh) + \nabla\varphi(x_0 - eh) - 2\nabla\varphi(x_0) = 0. \tag{220}$$

In addition, because $\delta\varphi$ is also stationary w.r.t. the unit vector e, we must have for any $\tau \perp e$,

$$h\left(\nabla_\tau\varphi(x_0 + eh) - \nabla_\tau\varphi(x_0 - eh)\right) := h\langle \tau, \nabla\varphi(x_0 + eh) - \nabla\varphi(x_0 - eh)\rangle = 0, \tag{221}$$

which means that $\nabla\varphi(x_0 + eh) - \nabla\varphi(x_0 - eh)$ is a multiple of e, i.e.,

$$\nabla\varphi(x_0 + eh) - \nabla\varphi(x_0 - eh) = 2\lambda e \tag{222}$$

for some $\lambda \in \mathbb{R}$. Combining (220) and (222), we see that

$$\begin{aligned}
0 &= [\nabla\varphi(x_0 + eh) - \nabla\varphi(x_0)] + [\nabla\varphi(x_0 - eh) - \nabla\varphi(x_0)], \\
2\lambda e &= [\nabla\varphi(x_0 + eh) - \nabla\varphi(x_0)] - [\nabla\varphi(x_0 - eh) - \nabla\varphi(x_0)].
\end{aligned}$$

Thus we must have

$$\begin{aligned}
\lambda e &= \nabla\varphi(x_0 + eh) - \nabla\varphi(x_0), \\
-\lambda e &= \nabla\varphi(x_0 - eh) - \nabla\varphi(x_0).
\end{aligned}$$

Since φ is convex, we have

$$\langle e, \lambda e\rangle \geqslant h\,\nabla^2\varphi(x_0)(e,e),$$

which gives

$$e^T[\lambda I_n - h\,\nabla^2\varphi(x_0)]e \geqslant 0.$$

So $\lambda \geqslant 0$. Note that

$$\begin{aligned}
\delta Q(\nabla\varphi(x_0)) &= Q(\nabla\varphi(x_0 + he)) + Q(\nabla\varphi(x_0 - he)) - 2Q(\nabla\varphi(x_0)) \\
&= Q(\nabla\varphi(x_0) + \lambda e) + Q(\nabla\varphi(x_0) - \lambda e) - 2Q(\nabla\varphi(x_0)) \\
&= 2\lambda^2 B(e,e),
\end{aligned}$$

which gives the "curvature" of $y = y(x) = \nabla\varphi(x)$. Moreover, since F is convex,

$$\begin{aligned}
\delta F(\nabla\varphi(x_0)) &= F(\nabla\varphi(x_0 + he)) + F(\nabla\varphi(x_0 - he)) - 2F(\nabla\varphi(x_0)) \\
&= F(\nabla\varphi(x_0) + \lambda e) + F(\nabla\varphi(x_0) - \lambda e) - 2F(\nabla\varphi(x_0)) \\
&\geqslant 0.
\end{aligned}$$

Now filling all pieces into (219) and using (218), we see that

$$0 \geqslant \delta\theta(x_0) = -\delta Q(x_0) + \delta Q(\nabla\varphi(x_0)) + \delta F(\nabla\varphi(x_0)) \geqslant -\delta Q(x_0) + \delta Q(\nabla\varphi(x_0))$$

and thus

$$2\lambda^2 B(e,e) \leqslant 2h^2 B(e,e).$$

Since $B(e,e) = Q(e) = \frac{1}{2}e^T Ae > 0$ and $\lambda \geqslant 0$, we have $0 \leqslant \lambda \leqslant h$. Recall the integral representation of $\delta\varphi$ in (217):

$$\delta\varphi(x_0) = \int_0^h \langle \nabla\varphi(x_0 + te) - \nabla\varphi(x_0 - te), e\rangle \,\mathrm{d}t. \tag{223}$$

It is easy to check that the integrand is non-decreasing in t. Indeed,

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi(x_0 + te) = \langle \nabla\varphi(x_0 + te),\, e\rangle, \quad \frac{\mathrm{d}}{\mathrm{d}t}\varphi(x_0 - te) = -\langle \nabla\varphi(x_0 - te),\, e\rangle, \tag{224}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\varphi(x_0 + te) = \langle \nabla^2\varphi(x_0 + te)e,\, e\rangle \geqslant 0, \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2}\varphi(x_0 - te) = \langle \nabla^2\varphi(x_0 - te)e,\, e\rangle \geqslant 0, \tag{225}$$

because φ is convex. So

$$\begin{aligned}\delta\varphi(x_0) &\leqslant \int_0^h \langle \nabla\varphi(x_0 + he) - \nabla\varphi(x_0 - he), e\rangle \,\mathrm{d}t \\ &= \int_0^h \langle 2\lambda e, e\rangle \,\mathrm{d}t \leqslant \int_0^h 2h \,\mathrm{d}t = 2h^2,\end{aligned}$$

i.e., we have shown that $\delta\varphi(x) \leqslant 2h^2$ uniformly in x. So all the eigenvalues of $\nabla^2\varphi$ are between 0 and 2. Thus,

$$|y(x_1) - y(x_2)| \leqslant 2|x_1 - x_2| \quad \text{for all } x_1, x_2 \in \mathbb{R}^n, \tag{226}$$

and we are off (210) by a factor of 2.

Step 3: bootstrap the priori estimate to conclude.

By the computation (223), (224), (225) in Step 2, we have seen that

$$\xi_t := \frac{\mathrm{d}}{\mathrm{d}t}\langle \nabla\varphi(x_0 + te) - \nabla\varphi(x_0 - te), e\rangle = \langle \nabla^2\varphi(x_0 + te)e,\, e\rangle + \langle \nabla^2\varphi(x_0 - te)e,\, e\rangle.$$

If $\nabla^2\varphi(x_0) \leqslant \alpha_0 I_n$ for some $1 < \alpha_0 \leqslant 2$ as a priori estimate, then $\xi_t \leqslant 2\alpha_0$ and

$$\begin{aligned}\delta\varphi(x_0) &\leqslant \int_0^h \min\{2h, 2\alpha_0 t\} \,\mathrm{d}t \\ &\leqslant \int_0^{h/\alpha_0} 2\alpha_0 t \,\mathrm{d}t + \int_{h/\alpha_0}^h 2h \,\mathrm{d}t \\ &= 2h^2 - \frac{h^2}{\alpha_0} = \left(2 - \frac{1}{\alpha_0}\right) h^2,\end{aligned}$$

i.e., we get a new bound

$$\delta\varphi \leqslant \alpha_1 h^2, \quad \text{where } \alpha_1 = 2 - \frac{1}{\alpha_0}.$$

Now we iterate to get $\delta\varphi \leqslant \alpha_k h^2$, where

$$\alpha_{k+1} = 2 - \frac{1}{\alpha_k} \quad \text{for } k = 0, 1, 2, \dots \quad \text{with } \alpha_0 = 2.$$

Then it is easy to check that the sequence $(\alpha_k) \downarrow 1$ as $k \to \infty$, and the limit α_∞ of this sequence is the fixed point of $\alpha = 2 - \frac{1}{\alpha}$. Solving the last equation for $1 \leqslant \alpha \leqslant 2$, we see that $\alpha_\infty = 1$ and the proof of the Caffarelli contraction theorem is complete. ∎

Proof of Lemma 4.14. Since φ is convex, $y = \nabla\varphi$ is a monotone map, i.e., $\langle x - x_0, y(x) - y(x_0)\rangle \geqslant 0$. For a unit vector $N \in \mathbb{R}^n$, we let

$$\Gamma(q, \theta, N) = \{q + t\alpha : q \in \mathbb{R}^n, \alpha \in \mathbb{R}^n, |\alpha| = 1, |\text{angle}(\alpha, N)| \leqslant \theta, t > 0\}$$

be the cone with vertex q, angle θ, and central axis N. Denote $y_0 = y(x_0)$. Then we claim that

$$\Gamma\left(y_0, \frac{\pi}{2} - \varepsilon, N\right) \cap B_R \subset y\left(\Gamma(x_0, \pi - \varepsilon, N)\right). \tag{227}$$

Indeed, consider $(x_0, y(x_0)), (x, y(x))$, and $s, t > 0$. If $y = y_0 + t\beta$ and $x = x_0 + s\alpha$ for some $|\alpha| = |\beta| = 1$, where $y_0 = y(x_0)$ and $y = y(x)$, then by the monotonicity of the map y we have

$$0 \leqslant \langle x - x_0, y(x) - y(x_0)\rangle = \langle s\alpha, t\beta\rangle = st\langle\alpha, \beta\rangle.$$

So $\langle\alpha, \beta\rangle \geqslant 0$, i.e., $|\text{angle}(\alpha, \beta)| \leqslant \pi/2$. Then

$$|\text{angle}(\alpha, N)| \leqslant |\text{angle}(\alpha, \beta)| + |\text{angle}(\beta, N)| \leqslant \frac{\pi}{2} + \frac{\pi}{2} - \varepsilon = \pi - \varepsilon,$$

which proves the claim (227). Now volume comparison gives

$$\text{Vol}\left(\Gamma\left(y_0, \frac{\pi}{2} - \varepsilon, N\right) \cap B_R\right) \leqslant \text{Vol}_f\left(\Gamma(x_0, \pi - \varepsilon, N)\right),$$

where $f = e^{-Q}$. If we take $N = \frac{x_0}{|x_0|}$, then

$$\text{Vol}_f\left(\Gamma(x_0, \pi - \varepsilon, N)\right) = \int_{\Gamma(x_0, \pi-\varepsilon, N)} e^{-Q(x)}\, \mathrm{d}x \to 0 \quad \text{as } |x_0| \to \infty$$

exponentially fast. But the only way that $\text{Vol}_f\left(\Gamma(x_0, \pi - \varepsilon, N)\right) \to 0$ if $y_0 = y(x_0) \to RN$ where $N = \frac{x_0}{|x_0|}$. This proves Lemma 4.14. ∎

4.3.2 Generalization via reverse heat flow

4.3.3 Equality case

5 Mean curvature flow

In this section, we turn to the *nonlinear* and *geometric* heat equation that describes the time evolution of submanifolds deformed by some vector field that minimizing the volume functional. Let $M := M^n \subset \mathbb{R}^N$ be an n-dimensional *closed* submanifold embedded in an ambient Euclidean space $\mathbb{R}^N$. Let $\mathbf{v}$ be a vector field in $\mathbb{R}^N$ and ∇ be the Euclidean covariant derivative (cf. the definition of covariant derivative in Appendix D.5). For $p \in M$, denote T_pM as the tangent space to M at p, and $T_p^\perp M$ as the orthogonal complement of T_pM. Let $(e_i)_{i=1}^n$ be an orthonormal frame for T_pM. The divergence of $\mathbf{v}$ on M is defined as

$$\operatorname{div}_M(\mathbf{v}) = \sum_{i=1}^n \langle e_i, \nabla_{e_i}\mathbf{v}\rangle =: \langle e_i, \nabla_{e_i}\mathbf{v}\rangle, \tag{228}$$

where we use the Einstein summation convention to omit the summation notation. This convention will be used throughout the rest of this section.

Definition 5.1 (Laplacian on manifold). Let $M^n \subset \mathbb{R}^N$ be a submanifold and $f : \mathbb{R}^N \to \mathbb{R}$. The *Laplacian* (or *Laplacian-Beltrami*) operator on M is defined as

$$\Delta_M f = \operatorname{div}_M(\nabla^\top f) = \langle e_i, \nabla_{e_i}\nabla^\top f\rangle, \tag{229}$$

where $\nabla^\top f$ is the tangential component of ∇f such that $\nabla f = \nabla^\top f + \nabla^\perp f$ and $\nabla^\perp f$ is the orthogonal component of ∇f.

Definition 5.2 (Second fundamental form). Let $M^n \subset \mathbb{R}^N$ be a submanifold. The *second fundamental form* of M is defined as

$$A : T_pM \times T_pM \to T_p^\perp M, \tag{230}$$

$$A(X, Y) = \nabla_X^\perp Y, \tag{231}$$

where X and Y are vector fields tangential to M when restricted to M.

Remark 5.3. Note that $\nabla_X Y$ can be understood as $\nabla_{\overline{X}}\overline{Y}$, where $\overline{X}$ and $\overline{Y}$ are smooth extensions of X and Y from M to $\mathbb{R}^N$. Moreover, since $\nabla_X Y - \nabla_Y X = [X, Y]$ and the Lie bracket $[X, Y]$ of X and Y is a tangential vector field, we have $\nabla_X^\top Y - \nabla_Y^\top X = 0$. So A is bilinear and symmetric. In tensor calculus language, A is said to be a symmetric $(0, 2)$-tensor taking values in the normal bundle. Finally, we have the following decomposition

$$\nabla_X Y = \nabla_X^\top Y + \nabla_X^\perp Y, \tag{232}$$

where $\nabla_X^\top Y$ is the induced connection on M and $\nabla_X^\perp Y$ is the second fundamental form of M. ■

Definition 5.4. Let $M^n \subset \mathbb{R}^N$ be a submanifold. The *mean curvature* of M is defined as

$$H = -A(e_i, e_i) = -\operatorname{tr}(A) = -\operatorname{tr}(\nabla_{e_i}^\perp e_i), \tag{233}$$

where $(e_i)_{i=1}^n$ is an orthonormal basis (ONB) for T_pM for $p \in M$.

Lemma 5.5 (Compute Δ_M on M of restriction of a function on $\mathbb{R}^N$). Let $M^n \subset \mathbb{R}^N$ be a submanifold and $(e_i)_{i=1}^n$ is an ONB for T_pM. For $f : \mathbb{R}^N \to \mathbb{R}$, we have

$$\Delta_M f = \nabla^2 f(e_i, e_i) - \langle H, \nabla f\rangle, \tag{234}$$

where $\nabla^2 f$ is the $\mathbb{R}^N$-Hessian of f.

Note that the Laplacian Δ_M on M is the trace of the Hessian on the tangent space with a correction term from the Euclidean subspace due to the second fundamental term.

Proof of Lemma 5.5. If $\mathbf{n}$ is a normal vector to T_pM, then by the Leibniz rule (cf. Appendix D.5),

$$0 = e_i\langle e_i, \mathbf{n}\rangle := \nabla_{e_i}\langle e_i, \mathbf{n}\rangle = \langle \nabla_{e_i} e_i, \mathbf{n}\rangle + \langle e_i, \nabla_{e_i}\mathbf{n}\rangle,$$

which implies that

$$\langle e_i, \nabla_{e_i}\mathbf{n}\rangle = -\langle \nabla_{e_i} e_i, \mathbf{n}\rangle = -\langle \nabla^{\perp}_{e_i} e_i, \mathbf{n}\rangle.$$

Then we compute

$$\begin{aligned}
\Delta_M f &= \operatorname{div}_M(\nabla^{\top} f) = \operatorname{div}_M(\nabla f) - \operatorname{div}_M(\nabla^{\perp} f) \\
&= \langle e_i, \nabla_{e_i}\nabla f\rangle - \langle e_i, \nabla_{e_i}\nabla^{\perp} f\rangle \\
&= \nabla^2 f(e_i, e_i) + \langle \nabla^{\perp}_{e_i} e_i, \nabla^{\perp} f\rangle \\
&= \nabla^2 f(e_i, e_i) + \langle \nabla^{\perp}_{e_i} e_i, \nabla f\rangle \\
&= \nabla^2 f(e_i, e_i) + \langle H, \nabla f\rangle,
\end{aligned}$$

because $\nabla^{\perp} f$ is normal. ∎

Throughout the rest of this section, we shall adopt the following notation. Let $f : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}$ be a function $f(x,t)$ of space and time. We use $\frac{\partial f}{\partial t}$ to denote the partial derivative of f w.r.t. t. If $x := x(t)$ is a function of t, we use $\partial_t f$ to denote the total time derivative of f, i.e.,

$$\partial_t f = \frac{\mathrm{d}}{\mathrm{d}t} f(x(t), t) = \langle \nabla f, x_t\rangle + \frac{\partial f}{\partial t}, \tag{235}$$

where $x_t = \frac{\partial x}{\partial t}$.

5.1 First variation of volume functional

Given an infinitely differentiable (i.e., smooth), compactly supported, normal vector field $\mathbf{v}$ on M, consider the one-parameter variation

$$M_{t,\mathbf{v}} = \{x + t\mathbf{v}(x) : x \in M\}, \tag{236}$$

which gives a curve $t \mapsto M_{t,\mathbf{v}}$ in the space of submanifolds with $M_{0,\mathbf{v}} = M$. To study the geometric flow of the one-parameter family submanifolds, we need first to compute the time derivative of volume, which is given by the first variation formula of volume.

Lemma 5.6 (First variation of volume functional). Let $M^n \subset \mathbb{R}^N$ be a submanifold and $\mathbf{v}$ be a field with compact support in M. The *first variation* formula of volume is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathrm{Vol}(M_{t,\mathbf{v}}) = \int_M \langle \mathbf{v}, H\rangle, \tag{237}$$

where H is the mean curvature vector in (233) and the integration $\int_M$ is with respect to the volume form dVol_M.

Proof of Lemma 5.6. Note that derivative of the volume element $\mu := \mu(t)\mathrm{dVol}(M_{t,\mathbf{v}})$ of $M_{t,\mathbf{v}}$ satisfies:

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mu(t) = \mathrm{div}_M(\mathbf{v})\mu(0).$$

Let $(e_i)_{i=1}^n$ be an ONB for M. Note that

$$\begin{aligned}
\mathrm{div}_M(\mathbf{v}) &= \mathrm{div}_M(\mathbf{v}^\perp) + \mathrm{div}_M(\mathbf{v}^\top) \\
&= \langle \nabla_{e_i}\mathbf{v}^\perp, e_i\rangle + \mathrm{div}_M(\mathbf{v}^\top) \\
&= -\langle \mathbf{v}^\perp, \nabla_{e_i} e_i\rangle + \mathrm{div}_M(\mathbf{v}^\top) \\
&= \langle \mathbf{v}^\perp, H\rangle + \mathrm{div}_M(\mathbf{v}^\top) \\
&= \langle \mathbf{v}, H\rangle + \mathrm{div}_M(\mathbf{v}^\top),
\end{aligned}$$

because $\langle \mathbf{v}^\top, H\rangle = 0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mu(t) = \Big(\langle \mathbf{v}, H\rangle + \mathrm{div}_M(\mathbf{v}^\top)\Big)\,\mu(0).$$

Integrating on M to get

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{Vol}(M_{t,\mathbf{v}}) = \int_M \langle \mathbf{v}, H\rangle \,\mathrm{dVol}_M + \int_M \mathrm{div}_M(\mathbf{v}^\top)\,\mathrm{dVol}_M.$$

Since $\mathbf{v}$ has compact support in M, $\int_M \mathrm{div}_M(\mathbf{v}^\top)\,\mathrm{dVol}_M = 0$. Thus we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{Vol}(M_{t,\mathbf{v}}) = \int_M \langle \mathbf{v}, H\rangle \,\mathrm{dVol}_M.$$

∎

Example 5.7 (n-sphere). Consider the n-sphere $M = \{x \in \mathbb{R}^{n+1} : |x| = r\}$ with radius r embedded in $\mathbb{R}^{n+1}$ for $n \geqslant 1$ and $\mathbf{v} = x/|x|$ is the unit normal vector field (since M is a hypersurface in $\mathbb{R}^{n+1}$). Consequently the one-parameter variation of M is given by

$$M_{t,\mathbf{v}} = \left\{x + s\frac{x}{|x|} : x \in M\right\} = \left\{\left(1 + \frac{t}{r}\right)x : |x| = r\right\}. \tag{238}$$

Then,

$$\mathrm{Vol}(M_{t,\mathbf{v}}) = \left(1 + \frac{t}{r}\right)^n \mathrm{Vol}(M) \tag{239}$$

and

$$\frac{\mathrm{Vol}(M_{t,\mathbf{v}}) - \mathrm{Vol}(M)}{t} = \frac{\left(1 + \frac{s}{r}\right)^n - 1}{t}\mathrm{Vol}(M). \tag{240}$$

Letting $t \to 0$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{Vol}(M_{t,\mathbf{v}}) = \frac{n}{r}\,\mathrm{Vol}(M). \tag{241}$$

On the other hand,

$$\langle H, \mathbf{v}\rangle = -\sum_{i=1}^n \langle \nabla_{e_i}^\perp e_i, \mathbf{v}\rangle = \sum_{i=1}^n \langle e_i, \nabla_{e_i}\mathbf{v}\rangle = \mathrm{div}_M(\mathbf{v}) = \mathrm{div}_M\left(\frac{x}{|x|}\right) = \frac{n}{r}, \tag{242}$$

which implies that

$$\int_M \langle H, \mathbf{v} \rangle = \frac{n}{r} \operatorname{Vol}(M). \tag{243}$$

Combining (241) and (243), we see that (237) indeed holds for n-sphere. ■

The first variation formula in Lemma 5.6 together with the Cauchy-Schwarz inequality imply that the volume functional is minimized at $\mathbf{v}^\perp = -H$. For smoothly evolving submanifolds $M_t^n = \{F(x,t) : x \in M^n, t \in \mathbb{R}\}$ for some function $F : M \times \mathbb{R} \to \mathbb{R}^N$, this means that $x_t^\perp = -H$ is the (negative) gradient flow of the volume functional, where $x_t = \frac{\partial x}{\partial t}$ and $x := x(t) = F(x,t)$ for $x \in M$.

Definition 5.8 (Mean curvature flow). Let $M_t^n \subset \mathbb{R}^N$ be submanifolds. $M_t := M_t^n$ is said to flow by the *mean curvature flow* (MCF) if

$$x_t^\perp = -H, \tag{244}$$

where $x_t^\perp$ is the normal component of the time derivative $x_t = \frac{\partial x}{\partial t}$ of the position vector $x := x(t) \in M_t$. If $n = 1$, then (244) is also called the *curve-shortening flow*.

In view of the first variation formula (237), we see that the mean curvature flow (244) is the negative gradient flow of volume.

Remark 5.9 (Reparametrization of MCF)*.* The MCF defined in (244) is often written as

$$x_t = -H. \tag{245}$$

Since a tangential vector field does not change the volume functional, the MCF defined by (244) and (245) only differ by a tangential diffeomorphism. Thus we have the same MCF (M_t) with different parameterization. In the sequel, we can either use (244) or (245) as our definition of the MCF. ■

Lemma 5.10 (Compute total time derivative of a function on the MCF)*.* Let $M_t := M_t^n \subset \mathbb{R}^N$ flow by the MCF and $f : \mathbb{R}^N \times I \to \mathbb{R}$ where $I \subset \mathbb{R}$. Then

$$\partial_t f = -\langle \nabla f, H \rangle + \frac{\partial f}{\partial t}, \tag{246}$$

Proof of Lemma 5.10. This lemma follows from (246) and the definition of the MCF (245). ■

Lemma 5.11*.* Let $M_t^n \subset \mathbb{R}^N$ flow by the MCF. Then we have

1. $(\partial_t - \Delta_{M_t})x_i = 0$ for all $i = 1, \ldots, N$.
2. $(\partial_t - \Delta_{M_t})|x|^2 = -2n$.
3. $(\partial_t - \Delta_{M_t})(|x|^2 + 2nt) = 0$.

Proof of Lemma 5.11. First we prove part 1. Let $\partial_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 is in the i-th position, i.e., ∂_i is the canonical Euclidean basis of $\mathbb{R}^N$. Then $\nabla x_i = \partial_i$ and $\nabla^2 f = 0$. So by Lemma 5.5, $\Delta_M x_i = -\langle H, \nabla f \rangle$. By Lemma 5.10, $\partial_t x_i = -\langle \nabla f, H \rangle$. Thus, we have $(\partial_t - \Delta_M)x_i = 0$.

To prove part 2, we note that $\nabla|x|^2 = 2x$ and $\nabla^2|x|^2 = 2I_N$. So by Lemma 5.5, $\Delta_M|x|^2 = \nabla^2|x|^2(e_i, e_i) - \langle H, \nabla f\rangle = 2n - \langle H, \nabla f\rangle$. By Lemma 5.10, $\partial_t|x|^2 = -\langle \nabla f, H\rangle$. Thus, we have $(\partial_t - \Delta_M)|x|^2 = -2n$.

To prove part 3, we note that $\nabla(|x|^2 + 2nt) = 2x$ and $\nabla^2(|x|^2 + 2nt) = 2I_N$. So by Lemma 5.5, $\Delta_M(|x|^2 + 2nt) = 2n - \langle H, \nabla f\rangle$. By Lemma 5.10, $\partial_t(|x|^2 + 2nt) = -\langle \nabla f, H\rangle + 2n$. Thus, we have $(\partial_t - \Delta_M)(|x|^2 + 2nt) = 0$. ■

Below we give several examples of the MCF.

Example 5.12 (Hyperplane). If M is an n-dimensional affine hyperplane, then $H = 0$, i.e., hyperplane is static solution of the MCF.

Example 5.13 (Evolving n-sphere). Using (242) in Example 5.7, we see that the mean curvature of an n-sphere of radius r equals to

$$H = \frac{n}{r}\frac{x}{|x|}.$$

By the definition $x_t^\perp = -H$ of the MCF in (244), the MCF on n-spheres is governed by the following ODE:

$$r'(t) = -\frac{n}{r(t)}, \tag{247}$$

where the right-hand side is the negative mean curvature of the n-sphere of radius $r(t)$. Since $(r^2)_t = 2r_t r = -2n$ for $r := r(t)$, the solution of the ODE (247) is given by

$$r(t) = \sqrt{r(0) - 2nt}, \tag{248}$$

which means that the solution of the MCF becomes extinct at a finite time $t = \frac{r(0)}{2t}$. In other words, when $t \uparrow \frac{r(0)}{2t}$, $H \to \infty$ so extinction point is a singularity of the flow.

Example 5.14 (Shrinking cylinder). Let

$$M_t^n = \mathbb{S}^k_{\sqrt{C-2kt}} \times \mathbb{R}^{n-k} \subset \mathbb{R}^N$$

be shrinking cylinders. Obviously, M_t^n contains shrinking spheres $\mathbb{S}^k_{\sqrt{C-2kt}}$ and a Euclidean factor $\mathbb{R}^{n-k}$. Using Example 5.13, the extinction point is $t = \frac{C}{2k}$ along the Euclidean factor $\mathbb{R}^{n-k}$.

Definition 5.15 (Minimal surface). A submanifold M is *minimal* if $H = 0$.

5.2 Parabolic maximum principle

Similarly as the linear heat equation in the Euclidean space (Section 2.2), the MCF also satisfies the parabolic maximum principle.

Theorem 5.16 (Parabolic maximum principle of the MCF). Let $M_t \subset \mathbb{R}^N$ be closed compact submanifolds and $(\partial_t - \Delta_{M_t})u = 0$. Then the maximum of u is attained at $t = 0$, i.e.,

$$\max_{t\geqslant 0, x\in M_t} u(x,t) = \max_{x\in M_0} u(x,0). \tag{249}$$

There are several important consequence of the parabolic maximum principle of the MCF. The first one is a convex hull property.

Corollary 5.17 (Convex hull property). If $M_t^n \subset \mathbb{R}^N$ is a compact MCF and

$$M_0 \subset \Omega := \{x \in \mathbb{R}^N : x_i < 0\}, \tag{250}$$

i.e., M_0 lies in a half-space Ω, then $M_t \subset \Omega$ for $t > 0$.

Proof of Corollary 5.17. Since M_t^n is an MCF, by Lemma 5.11 (part 1),

$$(\partial_t - \Delta_{M_t})x_i = 0.$$

By the parabolic maximum principle in Theorem 5.16, for each $t > 0$,

$$\max_{x \in M_t} x_i \leqslant \max_{x \in M_0} x_i < 0.$$

This means that $M_t \subset \Omega$ for all $t > 0$. ∎

The next important consequence is a comparison principle that gives the estimate on extinction time of the MCF.

Corollary 5.18 (Estimate on extinction time). If $M_t^n \subset \mathbb{R}^N$ is a compact MCF and $M_0 \subset B_r$, then $M_t \subset B_{\sqrt{r^2 - 2nt}}$ and M_t becomes extinct in time at most $\frac{r^2}{2n}$.

Remark 5.19. Corollary 5.18 implies that any closed submanifolds under the MCF becomes extinct in finite time. In particular, the extinction time estimate $\frac{r^2}{2n}$ given in Corollary 5.18 is sharp in view of Example 5.13, where the extinction time for evolving n-spheres of radius r at $t = 0$ is $\frac{r^2}{2n}$. ∎

Proof of Corollary 5.18. Since M_t^n is an MCF, by Lemma 5.11 (part 3),

$$(\partial_t - \Delta_{M_t})(|x|^2 + 2nt) = 0.$$

By the parabolic maximum principle in Theorem 5.16, for each $t > 0$,

$$\max_{M_t}(|x|^2 + 2nt) \leqslant \max_{M_0} |x|^2 \leqslant r^2.$$

This means that for $t > 0$,

$$\max_{M_t} |x| \leqslant \sqrt{r^2 - 2nt}.$$

Thus when $t = \frac{r^2}{2n}$, we have $0 \leqslant \max_{M_t} |x| \leqslant 0$, which implies $\max_{M_t} |x| = 0$. ∎

5.3 Minimal surface equation

Let $\Omega \subset \mathbb{R}^2$ and $u : \Omega \to \mathbb{R}$. Define the *graph* Γ_u associated with u is the set of points

$$\Gamma_u = \{(x, y, u(x, y)) : (x, y) \in \Omega\}. \tag{251}$$

We consider the *minimal graph problem*: given the boundary value $u(\partial\Omega)$, we want to find a u such that Γ_u has the least area. More specifically, we look at the minimal graph problem in $\mathbb{R}^3$. Assume the existence of Γ_u.

Theorem 5.20 (Minimal surface equation for graph in $\mathbb{R}^3$). Let $\Omega \subset \mathbb{R}^2$ and $u : \Omega \to \mathbb{R}$. Let Γ_u be the graph associated with u. Fix the boundary value of $u(\partial\Omega)$. If Γ_u has the least area, then we have the following *minimal surface equation*

$$\operatorname{div}\left(\frac{\eta \nabla u}{\sqrt{1+|\nabla u|^2}}\right) = 0 \tag{252}$$

holds in the distribution sense.

Proof of Theorem 5.20. We look for the variation $\eta : \Omega \to \mathbb{R}$ with $\eta(\partial\Omega) = 0$ such that $(u + t\eta)(\partial\Omega) = u(\partial\Omega)$ for all $t \in \mathbb{R}$. The area-minimality of Γ_u means that $\mathrm{Area}(t) := \mathrm{Area}(\Gamma_{u+t\eta})$ has a minimum at $t = 0$. Note that two tangent vectors on graph Γ_u can be taken as

$$\begin{aligned} \partial_x &\mapsto (1, 0, u_x), \\ \partial_y &\mapsto (0, 1, u_y). \end{aligned}$$

We remark that these two tangent vectors are linear independent, but may not be orthogonal. Then the area of Γ_u can be computed as:

$$\mathrm{Area}(\Gamma_u) = \int_\Omega |(1, 0, u_x) \times (0, 1, u_y)| \,\mathrm{d}x\,\mathrm{d}y,$$

where

$$u \times v = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} e_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} e_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} e_3,$$

$u = (u_1, u_2, u_3), v = (v_1, v_2, v_3)$, and $(e_i)_{i=1}^3$ are the canonical basis of $\mathbb{R}^3$. Now since

$$(1, 0, u_x) \times (0, 1, u_y) = -u_x e_1 - u_y e_2 + e_3$$

and

$$|(1, 0, u_x) \times (0, 1, u_y)| = \sqrt{u_x^2 + u_y^2 + 1},$$

we have for any function $u(x, y)$,

$$\mathrm{Area}(\Gamma_u) = \int_\Omega \sqrt{u_x^2 + u_y^2 + 1}\,\mathrm{d}x\,\mathrm{d}y = \int_\Omega \sqrt{|\nabla u|^2 + 1}\,\mathrm{d}x\,\mathrm{d}y.$$

Applying the last equality to $\Gamma_{u+t\eta}$, we have

$$\mathrm{Area}(t) = \int_\Omega \sqrt{1 + |\nabla(u + t\eta)|^2} = \int_\Omega \sqrt{1 + |\nabla u|^2 + 2t\langle \nabla u, \nabla \eta\rangle + t^2|\nabla \eta|^2},$$

which gives

$$\mathrm{Area}'(0) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{Area}(t) = \int_\Omega \frac{\langle \nabla u, \nabla \eta\rangle}{\sqrt{1+|\nabla u|^2}}.$$

Note that

$$\operatorname{div}\left(\frac{\eta \nabla u}{\sqrt{1+|\nabla u|^2}}\right) = \frac{\langle \nabla u, \nabla \eta\rangle}{\sqrt{1+|\nabla u|^2}} + \eta \operatorname{div}\left(\frac{\eta \nabla u}{\sqrt{1+|\nabla u|^2}}\right).$$

Since $\eta(\partial\Omega) = 0$, we have

$$\int_\Omega \operatorname{div}\left(\frac{\eta\nabla u}{\sqrt{1+|\nabla u|^2}}\right) = 0.$$

Then,

$$\text{Area}'(0) = -\int_\Omega \eta \operatorname{div}\left(\frac{\eta\nabla u}{\sqrt{1+|\nabla u|^2}}\right).$$

By the area-minimality (i.e., the first derivative test), we must have Γ_u, $\text{Area}'(0) = 0$. Since η is arbitrary, we conclude (252). ■

Theorem 5.21 (Bernstein). *If u is an entire solution of the minimal surface equation in $\mathbb{R}^2$, then $u(x,y) = ax + by + c$, i.e., Γ_u is a plane in $\mathbb{R}^3$.*

Remark 5.22. Bernstein's theorem is true up to $\mathbb{R}^7$, it is false in $\mathbb{R}^8$. ■

5.4 Huisken monotonicity

Define

$$H_b(x,t) = (-4\pi t)^{-N/2} \exp\left(\frac{|x|^2}{4t}\right) \qquad \text{for } x \in \mathbb{R}^N,\ t < 0. \tag{253}$$

It is easy to check that H_b defined in (253) is the fundamental solution of the *backward heat equation*:

$$(\partial_t + \Delta)H_b(x,t) = 0. \tag{254}$$

Let $u : \mathbb{R}^N \times (-\infty, 0) \to \mathbb{R}$ be a function $u(x,t)$ solving the (forward) heat equation $(\partial_t - \Delta)u = 0$ on negative time line. For $t < 0$, observe that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^N} uH_b = \int_{\mathbb{R}^N} u_t H_b + \int_{\mathbb{R}^N} u(H_b)_t = \int_{\mathbb{R}^N} (\Delta u)H_b - \int_{\mathbb{R}^N} u(\Delta H_b).$$

Then integration-by-parts give

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^N} uH_b = 0,$$

which means that $\int_{\mathbb{R}^N} uH_b$ is constant in time $t < 0$. In addition, by the reproducing property of the backward heat kernel,

$$\lim_{t\uparrow 0}\int_{\mathbb{R}^N} uH_b = u(0,0).$$

Thus we have

$$u(0,0) = \int_{\mathbb{R}^N} uH_b \qquad \text{for all } t < 0, \tag{255}$$

which is called *Watson's mean value property* of the backward heat equation.

Now we consider $M_t^n \subset \mathbb{R}^N$ flow by the MCF. Define $\Phi : \mathbb{R}^N \times (-\infty, 0) \to \mathbb{R}$ as:

$$\Phi(x,t) = (-4\pi t)^{-n/2} \exp\left(\frac{|x|^2}{4t}\right) \qquad \text{for } x \in \mathbb{R}^N,\ t < 0. \tag{256}$$

Note that Φ and H_b differ in exponent $(-4\pi t)^{-n/2}$ and $(-4\pi t)^{-N/2}$.

Theorem 5.23 (Huisken monotonicity). If $M_t := M_t^n \subset \mathbb{R}^N$ flow by the MCF, then for all $t < 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{M_t} \Phi = -\int_{M_t} \left| H + \frac{x^\perp}{2t} \right|^2 \Phi \leqslant 0, \tag{257}$$

with the equality attained if and only if $x_t^\perp = -H = \frac{x^\perp}{2t}$.

To prove the Huisken monotonicity, we need the following lemma.

Lemma 5.24. If $M_t := M_t^n \subset \mathbb{R}^N$ flow by the MCF, then

$$(\partial_t + \Delta_{M_t})\Phi = |H|^2\Phi - \left| H + \frac{x^\perp}{2t} \right|^2 \Phi. \tag{258}$$

Proof of Theorem 5.23. By the first variation formula of volume in Lemma 5.6,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{Vol}(M_{t,\mathbf{v}}) = \int_M \langle \mathbf{v}, H \rangle,$$

where $M_{t,\mathbf{v}} = \{x + t\mathbf{v}(x) : x \in M\}$. This implies for an MCF M_t^n that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{dVol}(M_t) = \langle x_t, H \rangle\, \mathrm{dVol}(M_t) = -\langle H, H \rangle\, \mathrm{dVol}(M_t) = -|H|^2\, \mathrm{dVol}(M_t).$$

Integrating and using the chain rule to get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{M_t} \Phi = \frac{\mathrm{d}}{\mathrm{d}t}\int \Phi\, \mathrm{dVol}(M_t) = \int_{M_t} \partial_t \Phi - \int_{M_t} |H|^2 \Phi.$$

Applying Lemma 5.24, we get

$$\int_{M_t} \partial_t \Phi = \int_{M_t} \left(|H|^2\Phi - \left| H + \frac{x^\perp}{2t} \right|^2 \Phi - \Delta_{M_t}\Phi \right).$$

Combining the last two displays, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{M_t} \Phi = -\int_{M_t} \left| H + \int_{M_t} \frac{x^\perp}{2t} \right|^2 \Phi - \int_{M_t} \Delta_{M_t}\Phi = -\int_{M_t} \left| H + \int_{M_t} \frac{x^\perp}{2t} \right|^2 \Phi,$$

where we used Stokes' theorem in the last equality. ■

Proof of Lemma 5.24. Denote ∇ and ∇^2 as the Euclidean covariant derivative and Hessian in $\mathbb{R}^N$, respectively. Denote ∇^{M_t} as the covariant derivative on M_t. For $t < 0$, direct calculation yields $\log \Phi(x,t) = -\frac{n}{2}\log(-4\pi t) + \frac{|x|^2}{4t}$, $\nabla \log \Phi = \frac{x}{2t}, \nabla^2 \log \Phi = \frac{1}{2t} I_N$, and $\frac{\partial}{\partial t}\log \Phi = -\frac{n}{2t} - \frac{|x|^2}{4t^2}$. For smooth functions $u(x,t)$ and $f : \mathbb{R} \to \mathbb{R}$, we compute

$$\begin{aligned}
(\partial_t + \Delta_{M_t}) f(u) &= f'(u)\partial_t u + \mathrm{div}_{M_t}(\nabla^{M_t} f(u)) \\
&= f'(u)\partial_t u + \mathrm{div}_{M_t}(f'(u)\nabla^{M_t} u) \\
&= f'(u)\partial_t u + f'(u)\mathrm{div}_{M_t}(\nabla^{M_t} u) + f''(u)\langle \nabla^{M_t} u, \nabla^{M_t} u \rangle \\
&= f'(u)(\partial_t u + \Delta_{M_t} u) + f''(u)|\nabla^{M_t} u|^2.
\end{aligned}$$

Applying $u = \log \Phi$ and $f(u) = e^u$, we see that $f(u) = f'(u) = f''(u) = e^u$ and

$$(\partial_t + \Delta_{M_t})\Phi = \Phi \left[(\partial_t + \Delta_{M_t}) \log \Phi + |\nabla^{M_t} \log \Phi|^2\right] \tag{259}$$

Next we compute the right-hand side of (259) with M_t flowing by the MCF. By Lemma 5.10,

$$\partial_t \log \Phi = -\langle \nabla \log \Phi, H\rangle + \frac{\partial}{\partial t} \log \Phi = -\langle \frac{x}{2t}, H\rangle - \frac{n}{2t} - \frac{|x|^2}{4t^2}.$$

By Lemma 5.5,

$$\Delta_{M_t} \log \Phi = \nabla^2 \log \Phi(e_i, e_i) - \langle H, \nabla \log \Phi\rangle = \frac{n}{2t} - \langle H, \frac{x}{2t}\rangle.$$

Adding the last two equations, we get

$$(\partial_t + \Delta_{M_t}) \log \Phi = -\frac{|x|^2}{4t^2} - 2\langle H, \frac{x}{2t}\rangle.$$

Moreover, note that

$$\nabla^{M_t} \log \Phi = \nabla^\top \log \Phi = \frac{x^\top}{2t}.$$

So we have

$$|\nabla^{M_t} \log \Phi|^2 = \left|\frac{x^\top}{2t}\right|^2 = \left|\frac{x}{2t}\right|^2 - \left|\frac{x^\perp}{2t}\right|^2 = \frac{|x|^2}{4t^2} - \frac{|x^\perp|^2}{4t^2}.$$

Putting all pieces together, we see that

$$(\partial_t + \Delta_{M_t}) \log \Phi + |\nabla^{M_t} \log \Phi|^2 = -\frac{|x|^2}{4t^2} - 2\langle H, \frac{x}{2t}\rangle + \frac{|x|^2}{4t^2} - \frac{|x^\perp|^2}{4t^2} = -2\langle H, \frac{x}{2t}\rangle - \frac{|x^\perp|^2}{4t^2}.$$

Filling the last equality into (259) and completing the square, we get

$$(\partial_t + \Delta_{M_t})\Phi = \Phi \left[-2\langle H, \frac{x}{2t}\rangle - \frac{|x^\perp|^2}{4t^2}\right] = \Phi \left[|H|^2 - \left|H + \frac{x^\perp}{2t}\right|^2\right].$$

■

5.5 Gaussian area and entropy

Definition 5.25 (Gaussian area). For a submanifold $M := M^n \subset \mathbb{R}^N$, the *Gaussian area* (or sometimes also called *Gaussian volume*) of M is defined as

$$F(M) = (4\pi)^{-n/2} \int_M e^{-|x|^2/4}. \tag{260}$$

Note that the normalization constant $(4\pi)^{-n/2}$ in Definition 5.25 makes $F = 1$ for any n-plane through the origin in $\mathbb{R}^N$.

Definition 5.26 (Entropy). For a submanifold $M := M^n \subset \mathbb{R}^N$, the *entropy* of M is defined as

$$\lambda(M) = \sup_{s>0, x_0 \in \mathbb{R}^N} F(sM + x_0). \tag{261}$$

Obviously, $\lambda(M) \geqslant 0$. In fact, $\lambda(M) \geqslant 1$ for any submanifold M.

Remark 5.27 (Invariance of entropy). By definition, the entropy λ in (261) is invariant under translation and scaling, i.e., for any $s > 0$ and $y_0 \in \mathbb{R}^N$, we have $\lambda(M) = \lambda(sM + y_0)$. In addition, λ is also invariant under rotations because the Gaussian weight in (260) satisfies $e^{-|x|^2/4} = e^{-|Qx|^2/4}$ for any $N \times N$ rotation matrix Q. ■

Theorem 5.28 (Entropy monotonicity). If $M_t := M_t^n \subset \mathbb{R}^N$ flows by the MCF, then $\lambda(M_t)$ is non-increasing in t. So $\lambda(M_t)$ is a Lyapunov function (i.e., a monotone quantity along the flow).

Proof of Theorem 5.28. ■

5.6 Shrinkers

A Multivariable calculus

A.1 Divergence theorem

Theorem A.1 (Divergence theorem on a bounded domain). Let $\Omega \subset \mathbb{R}^n$ and v is a vector field on Ω. Then

$$\int_\Omega \operatorname{div}(v) = \int_{\partial\Omega} \langle v, \mathbf{n} \rangle, \tag{262}$$

where $\mathbf{n}$ is the outer unit normal of Ω.

Corollary A.2. If $\eta : \mathbb{R}^n \to \mathbb{R}$ is a cutoff function such that $\eta \equiv 0$ on $\partial\Omega$, then

$$\int_\Omega \operatorname{div}(\eta v) = \int_{\partial\Omega} \langle \eta v, \mathbf{n} \rangle = 0, \tag{263}$$

where $\mathbf{n}$ is the outer unit normal of Ω.

Lemma A.3 (Integration-by-parts: no boundary term). Let v be a vector field on $\Omega \subset \mathbb{R}^n$ and $\eta : \mathbb{R}^n \to \mathbb{R}$ be a cutoff function such that $\eta \equiv 0$ on $\partial\Omega$. Then

$$\int_\Omega \eta \operatorname{div}(v) = -\int_\Omega \langle \nabla\eta, v \rangle. \tag{264}$$

Proof of Lemma A.3. By the chain rule,

$$\operatorname{div}(\eta v) = \langle \nabla\eta, v \rangle + \eta \operatorname{div}(v). \tag{265}$$

Applying Corollary A.2, we have

$$0 = \int_\Omega \operatorname{div}(\eta v) = \int_\Omega \langle \nabla\eta, v \rangle + \eta \operatorname{div}(v). \tag{266}$$

■

Theorem A.4 (Divergence theorem on manifolds).

$$\int_\Omega \operatorname{div}(v) \, \mathrm{dVol} = \int_{\partial\Omega} \langle v, \mathbf{n} \rangle \, \mathrm{d}\sigma, \tag{267}$$

where dVol is the volume element of Ω and $\mathrm{d}\sigma$ is the surface/boundary measure of $\partial\Omega$.

B Stochastic calculus

B.1 Itô's formula

Lemma B.1 (Itô's formula). Let $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a twice differentiable function and $(X_t)_{t\geqslant 0}$ is an n-dimensional Itô process

$$\mathrm{d}X_t = m(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}W_t, \tag{268}$$

where $(W_t)_{t\geqslant 0}$ is the standard Brownian motion in $\mathbb{R}^n$. Then,

$$\mathrm{d}f(X_t, t) = \left[\frac{\partial f}{\partial t} + \langle \nabla f, m\rangle + \frac{1}{2}\operatorname{tr}\left(\sigma^T \nabla f \sigma\right)\right]\mathrm{d}t + \langle \nabla f, \sigma\,\mathrm{d}W_t\rangle. \tag{269}$$

C Some functional inequalities

In this section, we present some useful functional inequalities.

C.1 Logarithmic Sobolev inequalities

A probability measure π is said to satisfy the logarithmic Sobolev inequality (LSI) if there exists a $\kappa > 0$ such that for any $\nu \in \mathcal{P}(\mathbb{R}^n)$ and $\pi \ll \nu$,

$$H(\nu\|\pi) \leqslant \frac{1}{2\kappa} I(\nu\|\pi). \tag{270}$$

We first discuss some Gaussian Sobolev inequalties. Let γ be the standard Gaussian measure on $\mathbb{R}^n$, i.e., $\gamma(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)$.

Lemma C.1 (Gaussian LSI: information-theoretic version). For any $\nu \in \mathcal{P}(\mathbb{R}^n)$ and $\gamma \ll \nu$,

$$H(\nu\|\gamma) \leqslant \frac{1}{2} I(\nu\|\gamma). \tag{271}$$

From Lemma C.1, we see that the standard Gaussian measure γ satisfies the LSI with $\kappa = 1$. The equality in (271) is achieved if ν is a translation of γ. To see this, take

$$\nu(x) = (2\pi)^{-n/2}\exp\Big(-\frac{|x-y|^2}{2}\Big), \quad y \in \mathbb{R}^n. \tag{272}$$

Then

$$\begin{aligned} H(\nu\|\gamma) &= \int \log\Big(\frac{\nu}{\gamma}\Big)\,\mathrm{d}\nu = \int -\frac{1}{2}(|x-y|^2 - |x|^2)\,\mathrm{d}\nu \\ &= \int \langle x, y\rangle\,\mathrm{d}\nu - \frac{1}{2}\int |y|^2\,\mathrm{d}\nu = |y|^2 - \frac{1}{2}|y|^2 = \frac{1}{2}|y|^2. \end{aligned} \tag{273}$$

On the other hand, since $\gamma \ll \nu$, we have

$$\rho(x) = \frac{\mathrm{d}\nu}{\mathrm{d}\gamma} = \exp\Big(-\frac{|x-y|^2}{2} + \frac{|x|^2}{2}\Big) = \exp(\langle x, y\rangle)\exp\Big(-\frac{|y|^2}{2}\Big) \tag{274}$$

and

$$\nabla\rho(x) = y\exp(\langle x, y\rangle)\,\exp\Big(-\frac{|y|^2}{2}\Big) = y\rho. \tag{275}$$

Then

$$I(\nu\|\gamma) = \int \frac{|y|^2\rho^2}{\rho}\,\mathrm{d}\gamma = \int |y|^2\rho\,\mathrm{d}\gamma = |y|^2. \tag{276}$$

Combining (273) and (276), we conclude $H(\nu\|\gamma) = \frac{1}{2}I(\nu\|\gamma)$. In fact, translation is the only case that is currently known to achieve the equality case. Hence we pose the following conjecture.

Conjecture C.2 (Characterization of the equality case in the Gaussian LSI)*.* The equality case $H(\nu\|\gamma) = \frac{1}{2}I(\nu\|\gamma)$ in the Gaussian LSI is achieved if and only if ν is a translation of γ, i.e., $\nu(x) = \gamma(x+y)$ for some $y \in \mathbb{R}^n$.

Lemma C.3 (Gaussian LSI: functional version)*.* Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function such that $f > 0$ and $\int f^2\,\mathrm{d}\gamma = 1$. Then

$$\int f^2 \log f\,\mathrm{d}\gamma \leqslant \int |\nabla f|^2\,\mathrm{d}\gamma. \tag{277}$$

Equivalently, we can write

$$\mathrm{Ent}_\gamma(f^2) \leqslant 2\int |\nabla f|^2\,\mathrm{d}\gamma, \tag{278}$$

where $\mathrm{Ent}_\gamma(g) = \int g\log g\,\mathrm{d}\gamma$ is the entropy of the probability density $g > 0$.

It is easy to see that Lemma C.3 and C.1 are equivalent. Let $g = f^2$. Then $\nabla g = 2f\nabla f$ and $\nabla f = \frac{1}{2\sqrt{g}}\nabla g$. Note that (277) is equivalent to

$$\int g\log g\,\mathrm{d}\gamma \leqslant 2\int \Big|\frac{1}{2\sqrt{g}}\nabla g\Big|^2\,\mathrm{d}\gamma = \frac{1}{2}\int \frac{|\nabla g|^2}{g}\,\mathrm{d}\gamma. \tag{279}$$

Since $g > 0$ and $\int g = 1$, we can define a probability measure μ such that $\frac{\mathrm{d}\mu}{\mathrm{d}\gamma} = g$. Then (279) can be written as

$$H(\mu\|\gamma) \leqslant \frac{1}{2}I(\mu\|\gamma). \tag{280}$$

Since g is arbitrary, the equivalence between Lemma C.3 and C.1 follows.

There are Euclidean LSIs and the following Lemma C.4 is a version with sharp constant.

Lemma C.4 (Euclidean LSI)*.* Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function such that $\int f^2\,\mathrm{d}x = 1$. Then

$$\int f^2\log f^2\,\mathrm{d}x \leqslant \frac{n}{2}\log\Big(\frac{2}{n\pi e}\int |\nabla f|^2\,\mathrm{d}x\Big). \tag{281}$$

Lemma C.5 (Sobolev inequality)*.* Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function with compact support. Then for all $n \geqslant 3$,

$$\int |f|^{\frac{2n}{n-2}} \leqslant C(n)\int |\nabla f|^2, \tag{282}$$

where

$$C(n) = \frac{1}{n(n-2)}\Big(\frac{\Gamma(n)}{\Gamma(n/2)}\Big)^{\frac{2}{n}} > 2. \tag{283}$$

Note that the Gaussian LSIs (Lemma C.1 and C.3) are dimension-free inequalities, while the Euclidean LSI (Lemma C.4) and Sobolev inequality (Lemma C.5) are not. The constants in (281) and (282) are both sharp. Indeed, direct computation gives the constant $\frac{2}{n\pi e}$ from (283). Let $m = nk$ and $k \to \infty$ and take a function $f : \mathbb{R}^{nk} \to \mathbb{R}$ such that $\int f^2 \, \mathrm{d}x = 1$. Then

$$C(m) = \frac{1}{\pi nk(nk-2)} \Big(\frac{\Gamma(nk)}{\Gamma(nk/2)} \Big)^{\frac{2}{nk}} \sim \frac{2^{\frac{1}{nk}}}{\pi n^2 k^2} \frac{2nk}{e} \sim \frac{2}{nk\pi e} = \frac{2}{m\pi e}, \tag{284}$$

where we used Stirling's approximation

$$(nk)! \sim \sqrt{2\pi nk} \Big(\frac{nk}{e} \Big)^{nk}. \tag{285}$$

C.2 Talagrand's transportation inequalities

D Riemannian geometry

In this section, we collect some basic facts and results about Riemannian geometry.

D.1 Smooth manifolds

Definition D.1 (Topological manifold). A topological space $(M, \mathcal{O})$ is said to be an n-dimensional *topological manifold* if for any $p \in M$, there is an open set $U \in \mathcal{O}$ containing p such that there exists a map $x : U \to x(U) \subset \mathbb{R}^n$ (equipped with the standard topology on $\mathbb{R}^n$) satisfying: (i) x is invertible, i.e., there is a map $x^{-1} : x(U) \to U$; (ii) x is continuous; (iii) x^{-1} is continuous. In other words, x is a *homeomorphism* between U and $x(U)$.

The pair (U, x) in Definition D.1 is called a *chart* and x is called the *chart map*. For $p \in U$ and $x(p) = (x^1(p), \dots, x^n(p))$, $x^i(p)$ is the i-th coordinate of $x(p)$. An *atlas* is a collection of charts $\mathcal{A} = \{(U_\alpha, x_\alpha) : \alpha \in A\}$ such that $M = \cup_{\alpha \in A} U_\alpha$.

For two charts (U, x) and (V, y) such that $U \cap V \neq \emptyset$, the map $y \circ x^{-1} : \mathbb{R}^n \supset x(U \cap V) \to y(U \cap V) \subset \mathbb{R}^n$ is said to be the *chart transition map*. We say (U, x) and (V, y) are *♣-compatible* if either $U \cap V = \emptyset$, or $U \cap V \neq \emptyset$ if the chart transition maps $y \circ x^{-1} : x(U \cap V) \to y(U \cap V)$ and $x \circ y^{-1} : y(U \cap V) \to x(U \cap V)$ has certain ♣-property (as a map from $\mathbb{R}^n$ to $\mathbb{R}^n$). For instance, the ♣-property can be $C^0(\mathbb{R}^n), C^1(\mathbb{R}^n), \dots, C^\infty(\mathbb{R}^n)$, and so on. An atlas $\mathcal{A}$ is ♣-compatible if the transition maps between any two charts in $\mathcal{A}$ have the ♣-property.

Definition D.2 (Smooth manifold). A *smooth manifold* $(M, \mathcal{O}, \mathcal{A})$ is a topological manifold $(M, \mathcal{O})$ equipped with a $C^\infty(\mathbb{R}^n)$-compatible atlas $\mathcal{A}$.

In these notes, we shall only consider smooth manifolds unless otherwise indicated.

Definition D.3 (Smooth functions on manifold). Let $(M, \mathcal{O}, \mathcal{A})$ be a smooth manifold. A function $f : M \to \mathbb{R}$ is said to be a *smooth map* (or C^∞-map) if $f \circ y^{-1} : \mathbb{R}^n \to \mathbb{R}$ is smooth for every chart y in the atlas $\mathcal{A}$.

According to Definition D.3, the coordinate functions $x^1, \dots, x^n$ of any chart (U, x) in a C^∞-compatible atlas are all smooth maps.

A map $\phi : M \to N$ is *smooth* if there is a chart (U, x) in M and a chart (V, y) in N such that $\phi(U) \subset V$ and the map $y \circ \phi \circ x^{-1} : \mathbb{R}^m \supset x^{-1}(U) \to y(V) \subset \mathbb{R}^n$ is C^∞.

Definition D.4 (Diffeomorphism). Let $(M, \mathcal{O}_M, \mathcal{A}_M)$ and $(N, \mathcal{O}_N, \mathcal{A}_N)$ are two smooth manifolds. We say that $(M, \mathcal{O}_M, \mathcal{A}_M)$ and $(N, \mathcal{O}_N, \mathcal{A}_N)$ are *diffeomorphic* if there exists a bijection $\phi : M \to N$ such that $\phi : M \to N$ and $\phi^{-1} : N \to M$ are both smooth maps.

D.2 Tensors

Let $(V, +, \cdot)$ be a vector space and $(V^*, \oplus, \odot)$ be the dual space, where

$$V^* = \{\phi : V \xrightarrow{\sim} \mathbb{R}\} =: \mathrm{Hom}(V, \mathbb{R}) \tag{286}$$

is the set of linear functionals on V. An element $\phi \in V^*$ is called a *covector*.

Definition D.5 (Tensor). Let $(V, +, \cdot)$ be a vector space and $r, s \in \mathbb{N}_0 := \{0, 1, \dots\}$. An *(r,s)-tensor* T over V is an $\mathbb{R}$-multi-linear map

$$T : \underbrace{V^* \times \cdots \times V^*}_{r \text{ times}} \times \underbrace{V \times \cdots \times V}_{s \text{ times}} \xrightarrow{\sim} \mathbb{R}. \tag{287}$$

By Definition D.5, a covector $\phi \in V^*$ is a $(0,1)$-tensor over V. If $\dim(V) < \infty$, then $V \cong (V^*)^*$ is an isomorphism so that $v \in V$ can be identified as a linear map $V^* \xrightarrow{\sim} \mathbb{R}$, which means that v is a $(1,0)$-tensor.

Let V be an n-dimensional vector space with an (arbitrarily chosen) basis $(e_1, \dots, e_n)$. Then the *dual basis* $(\varepsilon^1, \dots, \varepsilon^n)$ for V^* is uniquely determined by

$$\varepsilon^i(e_j) = \delta^i_j, \tag{288}$$

where $\delta^i_j = 1$ if $i = j$, and $\delta^i_j = 0$ if $i \neq j$.

Definition D.6 (Components of tensor). Let T be an (r,s)-tensor over an n-dimensional vector space V such that $n < \infty$. Let $(e_1, \dots, e_n)$ be a basis of V and $(\varepsilon^1, \dots, \varepsilon^n)$ be the dual basis of V^*. Then the *components* of T w.r.t. the chosen basis are defined as the $(r+s)^n$ real numbers (or sometimes called *coefficients*)

$$T^{i_1, \dots, i_r}{}_{j_1, \dots, j_s} = T(\varepsilon^{i_1}, \dots, \varepsilon^{i_r}, e_{j_1}, \dots, e_{j_s}) \tag{289}$$

for $i_1, \dots, i_r, j_1, \dots, j_s \in \{1, \dots, n\}$.

As an example, consider a $(1,1)$-tensor T with components given by $T^i{}_j = T(\varepsilon_i, e_j)$. Then for any $v \in V$ and $\phi \in V^*$, we can express

$$T(\phi, v) = T\Big(\sum_{i=1}^n \phi_i \varepsilon^i,\, \sum_{j=1}^n v^j e_j\Big) = \sum_{i=1}^n \sum_{j=1}^n \phi_i v^j T(\varepsilon^i,\, e_j) = \sum_{i=1}^n \sum_{j=1}^n \phi_i v^j T^i{}_j. \tag{290}$$

Thus components fully determine the tensor (given the basis). To avoid write too many sums, we typically use the *Einstein summation convention*:

$$T(\phi, v) = \phi_i v^j T^i{}_j, \tag{291}$$

where the repeated up-and-down indices i and j are summed over.

D.3 Tangent and cotangent spaces

Definition D.7 (Velocity). Let $(M, \mathcal{O}, \mathcal{A})$ be a smooth manifold and $\gamma : \mathbb{R} \to M$ be a curve (at least C^1). Let $p \in M$ and $\gamma(t_0) = p$ for some $t_0 = p$. The *velocity* of γ at p is the linear map defined as

$$\begin{aligned} v_{\gamma,p} \quad : \quad & C^\infty(M) \xrightarrow{\sim} \mathbb{R}, \\ & f \mapsto v_{\gamma,p}(f) = (f \circ \gamma)'(t_0). \end{aligned} \tag{292}$$

Note that $(C^\infty(M), \oplus, \odot)$ forms a vector space equipped with $(f \oplus g)(p) = f(p) + g(p)$ and $(\lambda \odot g)(p) = \lambda \cdot g(p)$. Note further that $(C^\infty(M), \oplus, \odot)$ is not only a vector space, it is also a *ring*. However, $(C^\infty(M), \oplus, \odot)$ is not a field!

Definition D.8 (Tangent vector space). Let $(M, \mathcal{O}, \mathcal{A})$ be a smooth manifold. For each $p \in M$,

$$T_pM := \Big\{ v_{\gamma,p} : \gamma \text{ is a smooth curve in } M \Big\} \tag{293}$$

is the *tangent space* to M at p.

Intuitively, we may understand the velocity $v_{\gamma,p}$ (i.e., a tangent vector in T_pM) as the directional derivative along the curve γ at p. The following Lemma D.9 confirms that the tangent space T_pM is indeed a vector space.

Lemma D.9. For each $p \in M$, T_pM is a vector space of linear maps from $C^\infty(M)$ to $\mathbb{R}$.

Proof of Lemma D.9. We shall define $\oplus$ and $\odot$ to make $(T_pM, \oplus, \odot)$ a vector space. For $f \in C^\infty(M)$, we define

$$\begin{aligned} \oplus \quad : \quad & T_pM \times T_pM \to \mathrm{Hom}(C^\infty(M), \mathbb{R}), \\ & (v_{\gamma,p} \oplus v_{\delta,p})(f) = v_{\gamma,p}(f) + v_{\delta,p}(f), \end{aligned} \tag{294}$$

and

$$\begin{aligned} \odot \quad : \quad & \mathbb{R} \times T_pM \to \mathrm{Hom}(C^\infty(M), \mathbb{R}), \\ & (s \odot v_{\gamma,p})(f) = s \cdot v_{\gamma,p}(f). \end{aligned} \tag{295}$$

We still need to show that there are smooth curves σ and τ such that $v_{\sigma,p} = v_{\gamma,p} \oplus v_{\delta,p}$ and $v_{\tau,p} = s \odot v_{\gamma,p}$. We first construct the curve τ via

$$\begin{aligned} \tau \quad : \quad & \mathbb{R} \to M, \\ & t \mapsto \tau(t) = \gamma(st + t_0) =: \gamma \circ \mu_s(t), \end{aligned} \tag{296}$$

where $\mu_s(t) = st + t_0$. Now $\tau(0) = \gamma(t_0) = p$ and by the chain rule,

$$\begin{aligned} v_{\tau,p}(f) = (f \circ \tau)'(0) &= (f \circ \gamma \circ \mu_s)'(0) = \mu_s'(0) \cdot (f \circ \gamma)'(\mu_s'(0)) \\ &= s \cdot (f \circ \gamma)'(t_0) = s \cdot v_{\gamma,p}(f) = (s \odot v_{\gamma,p})(f). \end{aligned} \tag{297}$$

Next choose a chart (U, x) such that $p \in U$, and define

$$\begin{aligned} \sigma_x \quad : \quad & \mathbb{R} \to M, \\ & t \mapsto \sigma_x(t) = x^{-1}((x \circ \gamma)(t_0 + t) + (x \circ \delta)(t_1 + t) - (x \circ \gamma)(t_0)), \end{aligned} \tag{298}$$

where $\gamma(t_0) = \delta(t_1) = p$. Then $\sigma_x(0) = p$ and by the chain rule,

$$\begin{aligned}
v_{\sigma_x,p}(f) &= (f \circ \sigma_x)'(0) = \big((f \circ x^{-1}) \circ (x \circ \sigma_x)\big)'(0) \\
&= \big((x \circ \sigma_x)^i\big)'(0) \cdot \big(\partial_i (f \circ x^{-1})\big)(x(\sigma_x(0))) \\
&= \Big(\big((x \circ \gamma)^i\big)'(t_0) + \big((x \circ \delta)^i\big)'(t_1)\Big) \cdot \big(\partial_i (f \circ x^{-1})\big)(x(p)) \\
&= \Big(\big((x \circ \gamma)^i\big)'(t_0)\Big) \cdot \big(\partial_i (f \circ x^{-1})\big)(x(p)) + \Big(\big((x \circ \delta)^i\big)'(t_1)\Big) \cdot \big(\partial_i (f \circ x^{-1})\big)(x(p)) \\
&= \big((f \circ x^{-1}) \circ (x \circ \gamma)\big)'(t_0) + \big((f \circ x^{-1}) \circ (x \circ \delta)\big)'(t_1) \\
&= (f \circ \gamma)'(t_0) + (f \circ \delta)'(t_1) = v_{\gamma,p}(f) + v_{\delta,p}(f) = (v_{\gamma,p} \oplus v_{\delta,p})(f),
\end{aligned} \tag{299}$$

which the last line does not depend on the choice of the chart (U, x). ■

Notation. In the proof of Lemma D.9, we have seen that for a curve $\gamma : \mathbb{R} \to M$ such that $\gamma(0) = p$, where (U, x) is a chart containing p,

$$v_{\gamma,p}(f) = \underbrace{\big((x \circ \gamma)^i\big)'(0)}_{=:\dot{\gamma}_x^i(0)} \cdot \underbrace{\big(\partial_i (f \circ x^{-1})\big)(x(p))}_{=:\left(\frac{\partial f}{\partial x^i}\right)_p}, \quad f : M \to \mathbb{R} \text{ smooth,} \tag{300}$$

where we reserve $\partial_i g$ as the j-th Euclidean partial derivative of the function $g : \mathbb{R}^n \to \mathbb{R}$. Thus we write the velocity vector $v_{\gamma,p} \in T_pM$ as a linear map from $C^\infty(M)$ to $\mathbb{R}$ defined as

$$v_{\gamma,p} = \dot{\gamma}_x^i(0) \Big(\frac{\partial}{\partial x^i}\Big)_p, \tag{301}$$

where $\left(\frac{\partial}{\partial x^i}\right)_p$ is the chart-induced basis of T_pM and $\dot{\gamma}_x^i(0)$ are the components of $v_{\gamma,p}$ w.r.t. the chart-induced basis. We emphasize that $\left(\frac{\partial f}{\partial x^i}\right)_p$ in (301) are *not* partial derivatives of f because it does not make sense to speak of partial derivatives for a function defined on a manifold where there is no *global* canonical basis such as in $\mathbb{R}^n$. Once we see $\left(\frac{\partial f}{\partial x^i}\right)_p$, we need always to translate back to its definition in (300).

Lemma D.10 (Chart-induced basis of tangent space)*.* Let $(M, \mathcal{O}, \mathcal{A})$ be a smooth manifold and (U, x) be a chart in $\mathcal{A}$ that contains p. Then

$$\Big(\frac{\partial}{\partial x^1}\Big)_p, \ldots, \Big(\frac{\partial}{\partial x^n}\Big)_p \tag{302}$$

form a basis of T_pU. In particular, $\dim(T_pM) = d = \dim(M)$, where $\dim(T_pM)$ is the vector space dimension of T_pM and $\dim(M)$ is the dimension of the topological manifold $(M, \mathcal{O})$.

Proof of Lemma D.10. We have shown in (301) that every $v_{\gamma,p} \in T_pU$ can be expressed as a linear combination of $\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p$. It remains to show that these vectors are linearly independent. Since $\mathcal{A}$ is a $C^\infty(\mathbb{R}^n)$-compatible atlas, $x^j : U \to \mathbb{R}$ is smooth. By (300),

$$\alpha^i \Big(\frac{\partial}{\partial x^i}\Big)_p (x^j) = \alpha^i \big(\partial_i (x^j \circ x^{-1})\big)(x(p)) = \alpha^i \delta_i^j = \alpha^j, \tag{303}$$

since $x^j \circ x^{-1} : \mathbb{R}^n \to \mathbb{R}$ such that $(x^j \circ x^{-1})(c_1, \ldots, c_n) = c_j$. If

$$0 = \alpha^i \Big(\frac{\partial}{\partial x^i}\Big)_p, \tag{304}$$

then $\alpha^j = 0$ for all $j = 1, \ldots, n$, which means that the chart-induced basis vectors in (302) are linearly independent. ■

Definition D.11 (Cotangent space). The *cotangent space* T_p^*M is the dual space of T_pM, i.e.,

$$T_p^*M = \{\phi : T_pM \xrightarrow{\sim} \mathbb{R}\} \tag{305}$$

contains the set of linear functionals on T_pM.

Example D.12 (Gradient is a covector in the cotangent space). Let $f \in C^\infty(M)$ and consider

$$\begin{aligned}(\mathrm{d}f)_p \quad &: \quad T_pM \xrightarrow{\sim} \mathbb{R},\\ &\quad\; X \mapsto (\mathrm{d}f)_p(X) = Xf, \quad \text{for } X \in T_pM.\end{aligned} \tag{306}$$

We call $(\mathrm{d}f)_p$ is the *gradient* of f at $p \in M$. Clearly $(\mathrm{d}f)_p \in T_p^*M$ (i.e., $(\mathrm{d}f)_p$ is a covector) and it is a $(0,1)$-tensor over the vector space T_pM. Thus the components of $(\mathrm{d}f)_p$ w.r.t. the chart-induced basis by (U, x) (cf. (289)) is given by

$$\big((\mathrm{d}f)_p\big)_j = (\mathrm{d}f)_p\Big(\Big(\frac{\partial}{\partial x^j}\Big)_p\Big) = \Big(\frac{\partial f}{\partial x^j}\Big)_p \quad \text{for } j = 1, \ldots, n. \tag{307}$$

The interpretation of gradient is as follows. The directional derivative Xf of f along a tangent vector X at p is the gradient $(\mathrm{d}f)_p$ evaluated at X. Thus gradient $(\mathrm{d}f)_p$ gives the information of directional derivatives of f along all possible tangent vectors at such point p.

Lemma D.13 (Dual basis of cotangent space). Let $(M, \mathcal{O}, \mathcal{A})$ be a smooth manifold and (U, x) be a chart in $\mathcal{A}$ that contains p, where $x^j : U \to \mathbb{R}$ is the j-th coordinate of x. Then

$$(\mathrm{d}x^1)_p, \ldots, (\mathrm{d}x^n)_p \tag{308}$$

form a dual basis of T_p^*M w.r.t. the basis $\big(\frac{\partial}{\partial x^1}\big)_p, \ldots, \big(\frac{\partial}{\partial x^n}\big)_p$ of T_pM.

Proof of Lemma D.13. The lemma follows from

$$(\mathrm{d}x^i)_p\Big(\frac{\partial}{\partial x^j}\Big)_p = \Big(\frac{\partial x^i}{\partial x^j}\Big)_p = \delta^i_j \tag{309}$$

and the dual basis definition in (288). ■

Now suppose we have two charts (U, x) and (V, y) such that $U \cap V \neq \emptyset$. Take a point $p \in U \cap V$ and a tangent vector $X \in T_pM$. Then we may express X in the two charts using different local coordinate systems:

$$X = X_x^i \Big(\frac{\partial}{\partial x^i}\Big)_p = X_y^j \Big(\frac{\partial}{\partial y^j}\Big)_p. \tag{310}$$

Let $f \in C^\infty(M)$. Applying the (multi-variable) chain rule, we compute

$$\begin{aligned}\Big(\frac{\partial}{\partial x^i}\Big)_p f = \partial_i(f \circ x^{-1})(x(p)) &= \partial_i\big((f \circ y^{-1}) \circ (y \circ x^{-1})\big)(x(p)) \\ &= \partial_i(y^j \circ x^{-1})(x(p)) \cdot \big(\partial_j(f \circ y^{-1})\big)(y(p)) \\ &= \Big(\frac{\partial y^j}{\partial x^i}\Big)_p \Big(\frac{\partial f}{\partial y^j}\Big)_p. \end{aligned} \tag{311}$$

Thus the last two displays imply that

$$X_x^i \Big(\frac{\partial y^j}{\partial x^i}\Big)_p \Big(\frac{\partial}{\partial y^j}\Big)_p = X_y^j \Big(\frac{\partial}{\partial y^j}\Big)_p. \tag{312}$$

Since $\Big(\frac{\partial}{\partial y^j}\Big)_p$ is a basis vector, we obtain the formula for the change of vector components under a change of chart given by

$$X_y^j = \Big(\frac{\partial y^j}{\partial x^i}\Big)_p X_x^i, \tag{313}$$

which is a linear map at the given point p. However, we should emphasize that the global chart transformation is nonlinear.

We can also derive a formula for change of covector components under a change of chart. Let $\omega \in T_p^*M$. Then we may write

$$\omega = \omega_{(x)i}(\mathrm{d}x^i)_p = \omega_{(y)j}(\mathrm{d}y^j)_p. \tag{314}$$

By similar computations in the tangent space, one can show that

$$\omega_{(y)j} = \Big(\frac{\partial x^i}{\partial y^j}\Big)_p \omega_{(x)i} \tag{315}$$

and the matrix $\Big(\frac{\partial x^i}{\partial y^j}\Big)_p$ is the inverse of the matrix $\Big(\frac{\partial y^j}{\partial x^i}\Big)_p$.

D.4 Tangent bundles and tensor fields

We have defined the tangent and cotangent spaces at a given point $p \in M$ in Section D.3. In this section, we consider the field extension of these concepts to the whole manifold based on the theory of bundles.

Definition D.14 (Bundle and fiber). A *bundle* is a triple (E, M, π) such that

$$\pi : E \to M, \tag{316}$$

where E is a smooth manifold ("total space"), M is a smooth manifold ("base space"), and π is a surjective smooth map ("projection map"). For $p \in M$, the *fiber* over p is the pre-image $\pi^{-1}(p)$ of $\{p\}$.

Consider a smooth n-dimensional manifold $(M, \mathcal{O}, \mathcal{A})$. Let

$$TM = \bigsqcup_{p \in M} T_p M \tag{317}$$

be a total space and $\pi : TM \to M$ be the surjective map defined via $X \mapsto p$ where p is the unique point in M such that $X \in T_p M$. We would like first to turn TM into a topological manifold such that π is continuous. We consider the coarsest topology:

$$\mathcal{O}_{TM} = \{\pi^{-1}(U) : U \in \mathcal{O}\}, \tag{318}$$

which is sometimes also referred as the *initial topology* w.r.t. π.

Next we need to construct a C^∞-atlas on TM from the C^∞-atlas on M. Let

$$\mathcal{A}_{TM} = \{(TU, \xi_x) : (U, x) \in \mathcal{A}\}, \tag{319}$$

where the chart map $\xi_x : TU \to \mathbb{R}^{2n}$ is defined as

$$X \mapsto \xi_x(X) = \Big(\underbrace{(x^1 \circ \pi)(X), \ldots, (x^n \circ \pi)(X)}_{(U,x)\text{-coordinates of base point } \pi(X)}, \underbrace{(\mathrm{d}x^1)_{\pi(X)}(X), \ldots, (\mathrm{d}x^n)_{\pi(X)}(X)}_{\text{components of } X \text{w.r.t. } (U,x)} \Big). \tag{320}$$

Note that

$$\begin{aligned} \xi_x^{-1} \ &: \ \xi_x(TU) \to TU, \\ &(a^1, \ldots, a^n, b^1, \ldots, b^n) \mapsto \Big(b^1 \Big(\frac{\partial}{\partial x^1}\Big)_{x^{-1}(a^1, \ldots, a^n)}, \cdots, b^n \Big(\frac{\partial}{\partial x^n}\Big)_{x^{-1}(a^1, \ldots, a^n)} \Big), \end{aligned} \tag{321}$$

where $x^{-1}(a^1, \ldots, a^n) = \pi(X)$ is the base point. Then we need to check that the atlas $\mathcal{A}_{TM}$ is smooth. Let (U, x) and (V, y) be two charts in $\mathcal{A}$ such that $U \cap V \neq \emptyset$. Combining (320) and (321), we can compute the chart transition map from $\mathbb{R}^{2n}$ to $\mathbb{R}^{2n}$ in $\mathcal{A}_{TM}$ as following:

$$\begin{aligned} &(\xi_y \circ \xi_x^{-1})(a^1, \ldots, a^n, b^1, \ldots, b^n) = \xi_y \Big(b^j \Big(\frac{\partial}{\partial x^j}\Big)_{x^{-1}(a^1, \ldots, a^n)} \Big) \\ = &\Big(\ldots, (y^i \circ \pi) \Big(b^j \Big(\frac{\partial}{\partial x^j}\Big)_{x^{-1}(a^1, \ldots, a^n)} \Big), \ldots, \ldots, (\mathrm{d}y^i)_{x^{-1}(a^1, \ldots, a^n)} \Big(b^j \Big(\frac{\partial}{\partial x^j}\Big)_{x^{-1}(a^1, \ldots, a^n)} \Big) \cdots \Big) \\ &= \Big(\ldots, (y^i \circ x^{-1})(a^1, \ldots, a^n), \ldots, \ldots, b^j \Big(\frac{\partial y^i}{\partial x^j}\Big)_{x^{-1}(a^1, \ldots, a^n)}, \ldots \Big). \end{aligned} \tag{322}$$

Note that $(y^i \circ x^{-1})(a^1, \ldots, a^n)$ is just the i-th coordinate of chart transition map of $\mathcal{A}$ and

$$\Big(\frac{\partial y^i}{\partial x^j}\Big)_{x^{-1}(a^1, \ldots, a^n)} = \partial_j (y^i \circ x^{-1})(x(x^{-1}(a^1, \ldots, a^n)))) = \partial_j (y^i \circ x^{-1})(a^1, \ldots, a^n)). \tag{323}$$

Since $y \circ x^{-1}$ is smooth, it follows that $\xi_y \circ \xi_x^{-1}$ is also smooth. Thus $(TM, \mathcal{O}_{TM}, \mathcal{A}_{TM})$ is a smooth manifold. Because π is a projection from TM to M, both of which are smooth manifolds, we conclude that π is a smooth map. Thus the triple (TM, M, π) has a bundle structure, which is referred as the tangent bundle.

Definition D.15 (Tangent bundle). For a smooth manifold $(M, \mathcal{O}, \mathcal{A})$, the triple (TM, M, π) is called the *tangent bundle*. For simplicity, we sometimes call TM the tangent bundle.

Why do we care about tangent bundles? The reason is that any smooth vector field on a smooth manifold M can be represented as a smooth section of the tangent bundle (TM, M, π).

Definition D.16 (Section). Let (E, M, π) be a bundle. A *section* σ of the bundle is a map $\chi : M \to E$ such that $\pi \circ \chi = \mathrm{id}_M$, where id_M is the identity map on M.

Recall that $(C^\infty(M), \oplus, \odot)$ is not only a vector space, it is also a *ring*. Here we slightly abuse the notation by denoting this ring as $(C^\infty(M), +, \cdot)$. Define

$$\Gamma(TM) = \{\chi : M \to TM : \chi \text{ is a smooth section}\}. \tag{324}$$

The key idea is that *one can think of* $\Gamma(TM)$ *as a collection of smooth vector fields on* M. Note that $\chi : M \to TM$ such that $p \mapsto \chi(p) \in T_pM$. Thus given a smooth map $f \in C^\infty(M)$, we can view $\chi f : M \to \mathbb{R}$ defined via $p \mapsto \chi(p)f \in \mathbb{R}$. In words, a smooth vector field $\chi \in \Gamma(TM)$ on M acting on a smooth function $f \in C^\infty(M)$ gives another smooth function $\chi f \in C^\infty(M)$.

Now we want to give more algebraic structures to the set $\Gamma(TM)$. Let $\chi, \widetilde{\chi} \in \Gamma(TM)$ be two smooth vector fields on M and $f, g \in C^\infty(M)$. Define

$$\begin{aligned}(\chi \oplus \widetilde{\chi})(f) &= \chi f + \widetilde{\chi} f, && (325)\\ (g \odot \chi)(f) &= g \cdot \chi(f). && (326)\end{aligned}$$

Then $(\Gamma(TM), \oplus, \odot)$ would be a vector space, if $C^\infty(M)$ is a field. However, $C^\infty(M)$ is only a ring, so this makes $(\Gamma(TM), \oplus, \odot)$ a $C^\infty(M)$*-module*. Informally, we call $\Gamma(TM)$ is "tangent field."

Similarly, one can consider the vector fields on the cotangent bundle T^*M and construct the section $\Gamma(T^*M)$ as a $C^\infty(M)$-module. In words, an element in $\Gamma(T^*M)$ takes a vector field in $\Gamma(TM)$ and it produces a smooth function in $C^\infty(M)$. Informally, we also call $\Gamma(T^*M)$ "cotangent field."

Definition D.17 (Tensor field). An (r, s)*-tensor field* T over $\Gamma(TM)$ is a $C^\infty(M)$-multi-linear map

$$T : \underbrace{\Gamma(T^*M) \times \cdots \times \Gamma(T^*M)}_{r \text{ times}} \times \underbrace{\Gamma(TM) \times \cdots \times \Gamma(TM)}_{s \text{ times}} \xrightarrow{\sim} C^\infty(M). \tag{327}$$

By convention, a $(0, 0)$-tensor field is a smooth function $C^\infty(M)$.

In the language of tensor field, an element of the cotangent field $\Gamma(T^*M)$ is a $(0, 1)$-tensor field, which takes a vector field in $\Gamma(TM)$ and produces a smooth function in $C^\infty(M)$. Below we give such an example.

Example D.18 (Gradient tensor field as a cotangent field). The $C^\infty(M)$-linear map

$$\begin{aligned}\mathrm{d}f \ : \ & \Gamma(TM) \xrightarrow{\sim} C^\infty(M),\\ & \chi \mapsto \mathrm{d}f(\chi) := \chi f, && (328)\end{aligned}$$

where as before $(\chi f)(p) = \chi(p)f$ gives the directional derivative of f along the vector $\chi(p)$ at p, is the $(0, 1)$-tensor field over the smooth vector fields $\Gamma(TM)$ on M. In other words, the gradient tensor field $\mathrm{d}f$ gives the gradients of f (on the whole manifold M) along all possible smooth vector fields χ, whereas we recall that the gradient $(\mathrm{d}f)_p$ gives the directional derivatives of f along all possible tangent vectors at a given point p.

D.5 Connections and curvatures

Let $X \in \Gamma(TM)$ be a smooth vector field on M. (Note that we slightly change the notation for a vector field from χ in Appendix D.4 to X.)

In this section, we wish to extend the notion of directional derivative $X : C^\infty(M) \to C^\infty(M)$ defined through $(Xf)(p) = X(p)f$ for $p \in M$ to the notion of connections on tensor fields (cf. Example D.12 and D.18), which allows us to define straight lines and eventually leads to curvatures.

Definition D.19 (Connection). A *connection* (sometimes also referred as *affine connection*) ∇ on a smooth manifold $(M, \mathcal{O}, \mathcal{A})$ is a map taking a pair of a vector (field) X on M and an (r, s)-tensor field T over $\Gamma(TM)$ and sending them to an (r, s)-tensor (field) $\nabla_X T$, satisfying:

1. $\nabla_X f = Xf$ for $f \in C^\infty(M)$ (i.e., f is a $(0, 0)$-tensor);
2. Additivity (in T): for (r, s)-tensor fields T and S,
$$\nabla_X(T + S) = \nabla_X T + \nabla_X S; \tag{329}$$
3. Leibniz rule (in T): for $\omega \in \Gamma(T^*M), Y \in \Gamma(TM)$, and $(1, 1)$-tensor field T,
$$\nabla_X(T(\omega, Y)) = (\nabla_X T)(\omega, Y) + T(\nabla_X \omega, Y) + T(\omega, \nabla_X Y); \tag{330}$$
4. $C^\infty(M)$-linearity (in X): for $f \in C^\infty(M)$,
$$\nabla_{fX+Z} T = f\nabla_X T + \nabla_Z T. \tag{331}$$

We say $\nabla_X T$ is the *covariant derivative* of T in the direction of X.

Remark D.20 (Remark on the Leibniz rule). First, the Leibniz rule (330) is equivalent to the tensor product form. Let T be a (p, q)-tensor field and S be an (r, s)-tensor field. The *tensor product* $T \otimes S$ of T and S is defined as the $(p + r, q + s)$-tensor field such that

$$\begin{aligned}&(T \otimes S)(\omega_1, \dots, \omega_{p+r}, X_1, \dots X_{q+s}) \\ = {} & T(\omega_1, \dots, \omega_p, X_1, \dots, X_q) \cdot S(\omega_{p+1}, \dots, \omega_{p+r}, X_{q+1}, \dots, X_{q+s}),\end{aligned} \tag{332}$$

for $\omega_1, \dots, \omega_{p+r} \in \Gamma(T^*M)$ and $X_1, \dots, X_{q+s} \in \Gamma(TM)$. Then the Leibniz rule (330) can be expressed in terms of the tensor product:

$$\nabla_X(T \otimes S) = (\nabla_X T) \otimes S + T \otimes (\nabla_X S). \tag{333}$$

Second, the Leibniz rule (330) can defined on higher-order tensors, which are useful to describe curvatures of smooth manifolds. For instance, for $(1, 2)$-tensor T, the Leibniz rule in T becomes

$$\nabla_X(T(\omega, Y, Z)) = (\nabla_X T)(\omega, Y, Z) + T(\nabla_X \omega, Y, Z) + T(\omega, \nabla_X Y, Z) + +T(\omega, Y, \nabla_X Z). \tag{334}$$

■

Now we can equip a smooth manifold with an additional connection structure. A *smooth manifold with connection* is a quadruple of structures $(M, \mathcal{O}, \mathcal{A}, \nabla)$. Intuitively, one can think of ∇_X is an extension of the directional derivative X and ∇ is an extension of the differential d, where both extensions are seen from smooth functions to tensor fields.

How many ways can we determine the connection ∇ on a smooth manifold $(M, \mathcal{O}, \mathcal{A})$? It turns out the degree of freedom (without putting extra structures) is high. Actually there are infinitely many connections can be given on a smooth manifold.

To see this, let X and Y be vector fields on M. Take a chart (U, x) and note that $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ are $(1,0)$-tensor fields (through an isomorphic identification). By the $C^\infty(M)$-linearity (331) and the Leibniz rule (333) (in the tensor product form), we may compute

$$\begin{aligned} \nabla_X Y &= \nabla_{X^i \frac{\partial}{\partial x^i}} \Big(Y^j \frac{\partial}{\partial x^j} \Big) = X^i \nabla_{\frac{\partial}{\partial x^i}} \Big(Y^j \frac{\partial}{\partial x^j} \Big) \\ &= X^i \Big(\nabla_{\frac{\partial}{\partial x^i}} Y^j \Big) \Big(\frac{\partial}{\partial x^j} \Big) + X^i Y^j \Big(\nabla_{\frac{\partial}{\partial x^i}} \Big(\frac{\partial}{\partial x^j} \Big) \Big) \\ &= X^i \Big(\frac{\partial Y^j}{\partial x^i} \Big) \frac{\partial}{\partial x^j} + X^i Y^j \underbrace{\nabla_{\frac{\partial}{\partial x^i}} \Big(\frac{\partial}{\partial x^j} \Big)}_{=: \Gamma^k{}_{ji} \frac{\partial}{\partial x^k}}, \end{aligned} \tag{335}$$

where $\Gamma^k{}_{ji}$ are the (chart-dependent) *connection coefficient functions* (on M) of ∇ w.r.t. (U, x).

Definition D.21 (Christoffel symbols). Given a smooth manifold $(M, \mathcal{O}, \mathcal{A})$ and a chart $(U, x) \in \mathcal{A}$, the *Christoffel symbols* $\Gamma^k{}_{ji} := \Gamma_{(x)}{}^k{}_{ji}$ are the connection coefficient functions defining a connection on the smooth manifold via

$$\begin{aligned} \Gamma^k{}_{ji} \quad : \quad & U \to \mathbb{R}, \\ & p \mapsto \Gamma^k{}_{ji}(p) := \Big(\mathrm{d}x^k \Big(\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} \Big) \Big)(p). \end{aligned} \tag{336}$$

In one chart (U, x), the Christoffel symbols, we can write the vector field in (335) as

$$\nabla_X Y = X^i \Big(\frac{\partial Y^j}{\partial x^i} \Big) \frac{\partial}{\partial x^j} + X^i Y^j \Gamma^k{}_{ji} \frac{\partial}{\partial x^k} = X^i \Big[\Big(\frac{\partial Y^j}{\partial x^i} \Big) \frac{\partial}{\partial x^j} + Y^j \Gamma^k{}_{ji} \frac{\partial}{\partial x^k} \Big], \tag{337}$$

or we can write down its components

$$(\nabla_X Y)^m = X(Y^m) + \Gamma^m{}_{ji} \, Y^j \, X^i, \tag{338}$$

where Y^m is the m-th component of Y and the products in this expression are all pointwise products of C^∞ functions. By the duality between basis and Leibniz rule, one can show that

$$(\nabla_X \omega)_m = X(\omega_m) - \Gamma^j{}_{mi} \, \omega_j \, X^i. \tag{339}$$

Based on (338) and (339), given an n-dimensional smooth manifold $(M, \mathcal{O}, \mathcal{A})$, we can determine a connection ∇ from the n^3-many Christoffel symbols $\Gamma^k{}_{ji}$, and we are left with a huge freedom for choosing $\Gamma^k{}_{ji}$ in order to determine a manifold with smooth connection $(M, \mathcal{O}, \mathcal{A}, \nabla)$.

A first step to nail down the number of connections is to require the torsion-free structure.

Definition D.22 (Torsion). Let $(M, \mathcal{O}, \mathcal{A}, \nabla)$ be a smooth manifold with connection. The *torsion* of ∇ is the $(1,2)$-tensor field

$$T(\omega, X, Y) = \omega(\nabla_X Y - \nabla_Y X - [X, Y]), \tag{340}$$

where the Lie bracket $[X, Y]$ is the vector field defined by

$$[X, Y]f = X(Yf) - Y(Xf), \quad f \in C^\infty(M). \tag{341}$$

The manifold $(M, \mathcal{O}, \mathcal{A}, \nabla)$ (or simply the connection ∇) is said to be *torsion-free* if $T \equiv 0$.

It is easy to check that torsion $T(\omega, X, Y)$ defined in (340) is indeed a $C^\infty(M)$-multi-linear map. Note that, for chart-induced basis,

$$[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = 0, \tag{342}$$

so that the components of torsion are given by $T^k{}_{ij} = \Gamma^k{}_{ji} - \Gamma^k{}_{ij}$. Thus, on a chart-induced basis, the connection ∇ is torsion-free if the Christoffel symbols are symmetric $\Gamma^k{}_{ji} = \Gamma^k{}_{ij}$.

In order to define a curvature, it is first instructive to think about how the straight lines look like in a curved smooth manifold. It is intuitively clear to speak about straight lines in Euclidean spaces. This leads to the notion of *autoparallely transported curves*.

Definition D.23 (Paralle transport). A vector field X on M is said to be *parallelly transported* along a smooth curve $\gamma : \mathbb{R} \to M$ if

$$\nabla_{v_\gamma} X = 0, \tag{343}$$

where v_γ denotes the velocity vector field along γ.

Definition D.24 (Autoparallely transported curve). A smooth curve $\gamma : \mathbb{R} \to M$ is said to be *autoparallely transported* if

$$\nabla_{v_\gamma} v_\gamma = 0, \tag{344}$$

where v_γ again denotes the velocity vector field along γ.

In words, the velocity vector field along an autoparallely transported curve is constant along the curve. We can view such curves have constant velocity, which mimics the "no-acceleration" situation of straight lines in $\mathbb{R}^n$.

Definition D.25 (Riemann curvature). The *Riemann curvature* of a connection ∇ is the $(1,3)$-tensor field

$$\mathcal{R}(\omega, Z, X, Y) = \omega(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z). \tag{345}$$

It is also easy to check that the Riemann curvature $\mathcal{R}(\omega, Z, X, Y)$ defined in (345) is indeed a $C^\infty(M)$-multi-linear map. In a chart-induced basis $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$, components of the Riemann curvature tensor can be expressed in terms of Christoffel symbols:

$$\begin{aligned} \mathcal{R}^a{}_{bcd} &= \mathcal{R}\Big(\mathrm{d}x^a, \frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c}, \frac{\partial}{\partial x^d}\Big) \\ &= \frac{\partial}{\partial x^c}\Gamma^a{}_{db} - \frac{\partial}{\partial x^d}\Gamma^a{}_{cb} + \Gamma^a{}_{cs}\Gamma^s{}_{db} - \Gamma^a{}_{ds}\Gamma^s{}_{cb}. \end{aligned} \tag{346}$$

Remark D.26 (Euclidean space with Cartesian coordinates). Consider the Euclidean space as a smooth manifold $(\mathbb{R}^n, \mathcal{O}, \mathcal{A})$ equipped with the standard topology (i.e., topology generated by open subsets in $\mathbb{R}^n$) and a smooth atlas $\mathcal{A}$. Under different atlas, we may have a Euclidean space with different coordinate systems such as Cartesian and polar coordinate systems. By default, if we assume the chart $(\mathbb{R}^n, \mathrm{id}_{\mathbb{R}^n}) \in \mathcal{A}$ and for such chart,

$$\Gamma^k{}_{ji} = \Big(\mathrm{d}x^k\Big(\nabla^e_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}\Big)\Big) = 0, \tag{347}$$

where $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ is the global Cartesian basis, then we call ∇^e the Euclidean connection and the resulting the manifold with connection $(\mathbb{R}^n, \mathcal{O}, \mathcal{A}, \nabla^e)$ the n-dimensional Euclidean space. Usually, we suppress the superscript e and simply write $\nabla = \nabla^e$.

D.6 Riemannian manifolds and geodesics

Now we consider a smooth manifold with a metric structure and construct a connection on such metric manifold. On a metric manifold, we can speak of metric quantities such as speed or length of curves. It turns out, with an additional metric-compatibility requirement (plus the torsion-free requirement), one can uniquely determine a connection on the metric manifold such that the straight lines are the same as length-minimizing curves.

Let $g : \Gamma(TM) \times \Gamma(TM) \to C^\infty(M)$ be a symmetric $(0, 2)$-tensor field (i.e., $g(X, Y) = g(Y, X)$ for all vector fields X and Y in $\Gamma(TM)$). We first define a bundle isomorphism called the *musical map*

$$\begin{aligned} \flat \ : \ & \Gamma(TM) \to \Gamma(T^*M), \\ & X \mapsto \flat(X) \text{ such that } \flat(X)(Y) = g(X, Y). \end{aligned} \tag{348}$$

By construction, the musical map $\flat := \flat_g$ depends on the metric tensor g.

For a vector field $X \in \Gamma(TM)$, $\flat(X) \in \Gamma(T^*M)$ is a $(0, 1)$-tensor with components given by

$$\begin{aligned} (\flat(X))_a &= \flat(X)\Big(\frac{\partial}{\partial x^a}\Big) = g\Big(X, \frac{\partial}{\partial x^a}\Big) \\ &= g\Big(X^m \frac{\partial}{\partial x^m}, \frac{\partial}{\partial x^a}\Big) = X^m g\Big(\frac{\partial}{\partial x^m}, \frac{\partial}{\partial x^a}\Big) = X^m g_{am}. \end{aligned} \tag{349}$$

where we used $C^\infty(M)$-multi-linearity of g.

Definition D.27 (Metric). A (non-degenerate) metric g on a smooth manifold $(M, \mathcal{O}, \mathcal{A})$ is a symmetric $(0, 2)$-tensor field such that the musical map $\flat$ is a C^∞-isomorphism (i.e., $\flat$ is invertible) between the tangent field $\Gamma(TM)$ and the cotangent field $\Gamma(T^*M)$.

Given a (symmetric and non-degenerate) metric g, one can define its inverse g^{-1} as the symmetric $(2, 0)$-tensor field by

$$\begin{aligned} g^{-1} \ : \ & \Gamma(T^*M) \times \Gamma(T^*M) \to C^\infty(M), \\ & (\omega, \sigma) \mapsto \omega(\flat^{-1}(\sigma)). \end{aligned} \tag{350}$$

For a covector field $\omega \in \Gamma(T^*M)$, $\flat^{-1}(\omega) \in \Gamma(TM)$ is a $(1, 0)$-tensor as a vector field (via an isomorphic identification) with components given by

$$\begin{aligned} (\flat^{-1}(\omega))^a &= \mathrm{d}x^a(\flat^{-1}(\omega)) = g^{-1}(\mathrm{d}x^a, \omega) \\ &= g^{-1}(\mathrm{d}x^a, \omega_m \mathrm{d}x^m) = \omega_m g^{-1}(\mathrm{d}x^a, \mathrm{d}x^m) = \omega_m g^{am}. \end{aligned} \tag{351}$$

where we used $C^\infty(M)$-multi-linearity of g^{-1} and we denote g^{am} as the components of g^{-1}.

Informally, a $(1,1)$-tensor field can be represented by a symmetric matrix (say in a chart), and we know that a symmetric matrix has an eigendecomposition with real eigenvalues. For a general (r,s)-metric tensor field g, it does not have an eigendecomposition. Rather, it only has a *signature*.

Definition D.28 (Riemannian metric). A metric is called *Riemannian* (or sometimes called *positive-definite*) if its signature is $(+,\dots,+)$. Metrics with all other signatures are called *pseudo-Riemannian* metric.

Definition D.29 (Riemannian manifold). A Riemannian manifold $(M,\mathcal{O},\mathcal{A},g)$ is a smooth manifold $(M,\mathcal{O},\mathcal{A})$ equipped with a Riemannian metric tensor g.

Because a Riemannian metric is positive-definite, it defines an inner product structure on the tangent field $\Gamma(TM)$. In particular, on a Riemannian metric manifold $(M,\mathcal{O},\mathcal{A},g)$, the *speed* $s(t)$ of a curve at $\gamma(t)$ is defined as

$$s(t) = \sqrt{\Big(g(v_\gamma, v_\gamma)\Big)_{\gamma(t)}}, \tag{352}$$

where v_γ is the velocity vector field along the curve (i.e., $v_{\gamma,t}$ is the velocity of curve γ at t).

Definition D.30 (Length). Let $\gamma : [0,1] \to M$ be a smooth curve. Then the *length of* γ is defined as

$$L(\gamma) = \int_0^1 s(t)\,\mathrm{d}t = \int_0^1 \sqrt{\Big(g(v_\gamma, v_\gamma)\Big)_{\gamma(t)}}\,\mathrm{d}t \tag{353}$$

Lemma D.31 (Length is preserved under reparametrization). Let $\gamma : [0,1] \to M$ be a smooth curve and $\sigma : [0,1] \to [0,1]$ be a smooth, bijective, and increasing function. Then

$$L(\gamma) = L(\gamma \circ \sigma). \tag{354}$$

Definition D.32 (Geodesic). A curve $\gamma : [0,1] \to M$ is called a *geodesic* on a Riemannian manifold $(M,\mathcal{O},\mathcal{A},g)$ if it is a *stationary* curve w.r.t. the length functional L in (353).

Recall that the signature of a Riemannian manifold is $(+,\dots,+)$. Thus, given two end points on M, a geodesic corresponding to the stationary point of L is the length-minimizing curve on the manifold. Nevertheless, we should note that geodesic does not always exist: consider two points $(+1,+1)$ and $(-1,-1)$ in $\mathbb{R}^2 \setminus \{(0,0)\}$ with the Euclidean metric on $\mathbb{R}^2$. This problem can be fixed by considering the *admissible curve*, which is a piecewise smooth curve segment. Setting

$$d(p,q) = \inf\big\{L(\gamma) : \gamma : [0,1] \to M \text{ is admissible}, \gamma(0) = p, \gamma(1) = q\big\}, \tag{355}$$

then (M,d) is a metric space [8, Chapter 2] (sometimes referred as a *length space*). If (M,d) is a complete metric space, then it is a *geodesically complete manifold/space* or simply *geodesic space*.

Theorem D.33 (Levi-Civita connection). On a Riemannian manifold $(M,\mathcal{O},\mathcal{A},g)$, there is a unique connection ∇ that is torsion-free $T = 0$ and metric-compatible $\nabla g = 0$. This connection is called the *Levi-Civita connection* (or sometimes called the *Riemiannian* connection).

For any vector fields X, Y, Z in the tangent bundle TM, the Leibniz rule (330) reads

$$\nabla_X g(Y, Z) = (\nabla_X g)(Y, Z) + g(\nabla_X Y, Z) + g(Y, \nabla_X Z). \tag{356}$$

Since $g(Y, Z) \in C^\infty(M)$ and $\nabla_X g = 0$, metric compatibility can be equivalently expressed as

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z). \tag{357}$$

From now on, on a Riemannian manifold $(M, \mathcal{O}, \mathcal{A}, g)$, the connection endowed will always be the Levi-Civita connection (without mentioning further). In such case, the metric compatibility implies that a straight line between two points on the manifold is the same as a geodesic that minimizes the length functional L in (353). That is, once we work on a Riemannian manifold, the Riemman curvature tensor will be (implicitly) induced from the Levi-Civita connection. However, we should highlight that the Riemann curvature tensor can be defined through any connection without referring to the Riemannian metric.

Lemma D.34 (Existence and uniqueness of geodesics). Let $(M, \mathcal{O}, \mathcal{A}, g)$ be a Riemannian manifold. For every point $p \in M$ and $v \in T_pM$, there is a unique maximal geodesic $\gamma := \gamma_v : I \to M$ with $\gamma(0) = p$ and $\gamma'(p) = v$, defined on some open interval I containing 0.

Proof of Lemma D.34 can be found in Theorem 4.27 and Corollary 4.28 in [8].

Definition D.35 (Exponential map). Let $(M, \mathcal{O}, \mathcal{A}, g)$ be a Riemannian manifold and $p \in M$. The *exponential map* $\exp_p : T_pM \to M$ is defined by

$$\exp_p(v) = \gamma_v(1), \tag{358}$$

where γ_v is the unique maximal geodesic defined on an open interval containing $[0, 1]$ (cf. Lemma D.34).

Note that for a smooth manifold $(M, \mathcal{O}, \mathcal{A})$, without additional connection ∇ or metric g structures, we cannot speak of the manifold shape. Given two end points, the straight line determined by the autoparallely transported curves coincides to the geodesic determined by the metric g if coefficient functions of the Levi-Civita connection satisfy that

$$\Gamma^i{}_{jk} = \frac{1}{2} g^{iq}\Big(\frac{\partial g_{kq}}{\partial x^j} + \frac{\partial g_{jq}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^q}\Big), \tag{359}$$

which comes from the *geodesic equation*.

Example D.36 (Round metric on 2-sphere). Consider the 2-sphere manifold $(\mathbb{S}^2, \mathcal{O}, \mathcal{A})$ and a chart map $x(p) := (x^1(p), x^2(p)) = (\theta, \varphi)$ such that $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$. We now determine the shape of the smooth manifold 2-sphere by the *round metric*. Under the chart map x, we may equip the 2-sphere smooth manifold $(\mathbb{S}^2, \mathcal{O}, \mathcal{A})$ with the round metric whose components are given by

$$\big(g_{ij}(x^{-1}(\theta, \varphi))\big)_{i,j=1,2} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2(\theta) \end{pmatrix}. \tag{360}$$

It is straightforward to check that the metric g in (360) corresponds to a connection ∇^{round} with the connection coefficients in (359) given by

$$\Gamma^1{}_{22}(x^{-1}(\theta, \varphi)) = -\sin(\theta)\cos(\theta), \quad \Gamma^1{}_{12} = \Gamma^1{}_{21} = \cot(\theta), \tag{361}$$

and all other 5 Christoffel symbols are all zeros. Consider the equator curve γ of the round 2-sphere parameterized by

$$\theta(t) := (\theta\circ\gamma)(t) = (x^1\circ\gamma)(t) = \frac{\pi}{2}, \tag{362}$$

$$\varphi(t) := (\varphi\circ\gamma)(t) = (x^2\circ\gamma)(t) = 2\pi t^3. \tag{363}$$

Obviously, $\theta'(t) = 0$ and $\varphi'(t) = 6\pi t^2$. Then the length of γ can be computed by using the components of the round metric tensor

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{g_{ij}(x^{-1}(\theta(t),\varphi(t)))(x^i\circ\gamma)'(t)(x^j\circ\gamma)'(t)}\,\mathrm{d}t \\ &= \int_0^1 \sqrt{1\cdot 0 + \sin^2(\pi/2)\cdot 36\pi^2 t^4}\,\mathrm{d}t & (364) \\ &= 6\pi\int_0^1 t^2\,\mathrm{d}t = 2\pi, & (365) \end{aligned}$$

which is the same as the length of the equator curve under the constant speed parametrizaiton $\varphi(t) = 2\pi t$. In the special case of round sphere, this calculation verifies the length maintenance under reparametrization stated in Lemma D.31. ■

From the metric tensor g, we can define more curvature tensors on a Riemannian manifold via *contraction*.

Definition D.37 (Curvature tensors)*.* Let $(M,\mathcal{O},\mathcal{A},g)$ be a Riemannian manifold.

1. The *Riemann-Christoffel curvature* is a $(0,4)$-tensor defined by

$$R_{abcd} = g_{am}\,\mathcal{R}^m{}_{bcd}. \tag{366}$$

2. The *Ricci curvature* is a $(0,2)$-tensor defined by

$$R_{ab} = \mathcal{R}^m{}_{amb}. \tag{367}$$

3. The *scalar curvature* is defined by

$$R = (g^{-1})^{ab}\,R_{ab}. \tag{368}$$

4. The *Einstein curvature* is a $(0,2)$-tensor defined by

$$G_{ab} = R_{ab} - \frac{1}{2}R\,g_{ab}. \tag{369}$$

D.7 Volume forms

Consider the problem of integrating functions over a smooth manifold $(M,\mathcal{O},\mathcal{A})$.

Definition D.38 (Volume form)*.* On an n-dimensional smooth manifold $(M,\mathcal{O},\mathcal{A})$, a $(0,n)$-tensor field Ω (over $\Gamma(TM)$) is called a *volume form* if

1. Non-vanishing: $\Omega \neq 0$ on M;

2. Totally anti-symmetric: $\Omega(\dots, X_i, \dots, X_j, \dots) = -\Omega(\dots, X_j, \dots, X_i, \dots)$ for all $X_i, X_j \in \Gamma(TM)$ and $i, j = 1, \dots, n$.

We can construct a (natural) volume form from a Riemannian manifold. Let $(M, \mathcal{O}, \mathcal{A}, g)$ be a Riemannian manifold. Take an arbitrary chart (U, x) and we let the components of the tensor field Ω be given by

$$\Omega_{i_1,\dots,i_n} = \sqrt{g_{ij}}\, \epsilon_{i_1,\dots,i_n}, \tag{370}$$

where $(i_1, \dots, i_n)$ is a permutation of $(1, \dots, n)$ such that $\epsilon_{1,\dots,n} = 1$ and $\epsilon_{i_1,\dots,i_n} = \epsilon_{[i_1,\dots,i_n]}$ for any anti-symmetric bracket $[\dots]$. The $\epsilon_{i_1,\dots,i_n}$'s are called the *Levi-Civita symbols*. One can check that $\Omega_{i_1,\dots,i_n}$ is well-defined if and only if for any pair of charts (U, x) and (U, y),

$$\det\left(\frac{\partial y}{\partial x}\right) = \det(\partial_i(y^j \circ x^{-1})) > 0. \tag{371}$$

Condition (371) means that the chart transition maps are *oriented*.

Definition D.39 (Oriented atlas). Let $\mathcal{A}$ be a smooth atlas. Then $\mathcal{A}^\uparrow \subset \mathcal{A}$ is called a (positively) *oriented sub-atlas* of $\mathcal{A}$ if for any two charts $(U, x), (V, y) \in \mathcal{A}^\uparrow$, the chart transition maps $y \circ x^{-1}$ and $x \circ y^{-1}$ are oriented:

$$\det\left(\frac{\partial y}{\partial x}\right) > 0 \quad \text{or} \quad \det\left(\frac{\partial x}{\partial y}\right) > 0 \quad \text{on } U \cap V. \tag{372}$$

Let $(M, \mathcal{O}, \mathcal{A}^\uparrow)$ be a smooth oriented manifold and $(U, x) \in \mathcal{A}^\uparrow$ be an oriented chart. Given a volume form Ω, we can define a *scalar density* on M by

$$\omega(p) = \Omega_{i_1,\dots,i_n}(p)\, \epsilon^{i_1,\dots,i_n} \quad \text{for } p \in M, \tag{373}$$

where $\epsilon^{i_1,\dots,i_n} = \epsilon_{i_1,\dots,i_n}$. Note that $\Omega_{i_1,\dots,i_n}$ is chart-dependent, so is ω. Moreover, the change of scalar density under a change of chart is given by

$$\omega_{(y)} = \det\left(\frac{\partial x}{\partial y}\right) \omega_{(x)} \tag{374}$$

for any two charts $(U, x), (U, y) \in \mathcal{A}^\uparrow$.

Let $(M, \mathcal{O}, \mathcal{A}^\uparrow, g)$ be an oriented metric manifold. On one chart domain (U, x), we can define the integration of a function $f : U \to \mathbb{R}$ as

$$\int_U f := \int_{x(U)} \sqrt{\det(g_{ij})(x^{-1}(\alpha))}(f \circ x^{-1})(\alpha)\, \mathrm{d}\alpha. \tag{375}$$

One can check that (375) is well-defined and does not depend on the chart map on the same chart domain U. To extend the integration to the whole manifold, we would then need *partition of unity*. Let $(U_i, x_i) \in \mathcal{A}^\uparrow$ and $\varrho_i : U_i \to \mathbb{R}, i = 1, \dots, N$, be a finite collection of continuous functions such that for any $p \in M$, we have $\sum_i \varrho_i(p) = 1$ where the sum is taken over i such that $p \in U_i$. Then we define the integration of a function $f : M \to \mathbb{R}$ as

$$\int_M f = \sum_{i=1}^N \int_{U_i} (\varrho_i f). \tag{376}$$

Combining (375) and (376), we can define the (natural) volume form of a Riemannian manifold as following.

Definition D.40 (Volume form of oriented Riemannian manifold). Let $(M, \mathcal{O}, \mathcal{A}^{\uparrow}, g)$ be an oriented Riemmanian manifold and $(U, x) \in \mathcal{A}^{\uparrow}$ be an oriented chart. The *volume form* of the manifold M in an oriented chart (i.e., local coordinates) is defined as

$$\mathrm{dVol} = \sqrt{\det(g)}\, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n, \tag{377}$$

where $\wedge$ denotes the wedge product. The integration of a function $f : M \to \mathbb{R}$ is defined as

$$\int_M f := \int_M f(p)\, \mathrm{dVol}(p). \tag{378}$$

Acknowledgement

This lecture note was initially carried out when the author was visiting the Institute for Data, System, and Society (IDSS) at Massachusetts Institute of Technology and gradually expanded while the author was teaching a topics course “Geometric Flows” at the University of Illinois at Urbana-Champaign. The author would like to thank Sinho Chewi, Tobias Colding, Thibaut Le Gouic, Philippe Rigollet, Austin Stromme, Yun Yang for their helpful comments and feedbacks. This work is supported in part by an NSF CAREER award DMS-1752614 and a Simons Fellowship in Mathematics.

References

[1] David Alonso-Gutiérrez and Jesús Bastero. *Approaching the Kannan-Lovász-Simonovits and Variance Conjectures*. Springer, Cham, 2015.

[2] Francois Bolley, Ivan Gentil, and Arnaud Guillin. Convergence to equilibrium in wasserstein distance for fokker-planck equations. *Journal of Functional Analysis*, 263(8):2430 – 2457, 2012.

[3] Silouanos Brazitikos, Apostolos Giannopoulos, Petros Valettas, and Beatrice-Helen Vritsiou. *Geometry of Isotropic Convex Bodies*, volume 196 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2014.

[4] Luis A. Caffarelli. Monotonicity Properties of Optimal Transportation and the FKG and Related Inequalities. *Commun. Math. Phys.*, 214:547–563, 2000.

[5] Ivan Gentil, Christian Léonard, Luigia Ripani, and Luca Tamanini. An entropic interpolation proof of the hwi inequality. *Stochastic Processes and their Applications*, 2019.

[6] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[7] Young-Heon Kim and Emanuel Milman. A generalization of Caffarelli's contraction theorem via (reverse) heat flow. *Mathematische Annalen*, 354:827–862, 2012.

[8] John Lee. *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer International Publishing, 2 edition, 2018.

[9] Y. T. Lee and S. S. Vempala. Eldan's stochastic localization and the kls hyperplane conjecture: An improved lower bound for expansion. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 998–1007, Oct 2017.

[10] F. Otto and C. Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173(2):361 – 400, 2000.

[11] Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport: With Applications to Data Science*. Now Foundations and Trends, 2019.

[12] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians - Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser, 2015.

[13] Ramon van Handel. Probability in high dimension. APC 550 Lecture Notes, Princeton University, 2016.