Information Geometry (2024) 7:59–95
https://doi.org/10.1007/s41884-023-00124-x

RESEARCH PAPER

Diffusion hypercontractivity via generalized density manifold

Wuchen Li[1]

Received: 4 February 2020 / Revised: 12 October 2023 / Accepted: 29 October 2023 /
Published online: 27 November 2023

Abstract

We prove a one-parameter family of diffusion hypercontractivity from a class of drift-diffusion processes. We next derive the related log–Sobolev, Poincare, and Talagrand inequalities. The derivation is based on the calculation of Hessian operators along generalized gradient flows in Dolbeault–Nazaret–Savare metric spaces (Dolbeault et al., Calc Var Partial Differ 2:193–231, 2010). In this direction, a mean-field type Bakry–Emery iterative calculus is presented. In particular, an inequality among Pearson divergence (P), negative Sobolev metric (H^{-1}), and generalized Fisher information functional (I), named $PH^{-1}I$ inequality, is presented.

Keywords Information theory · Mean-field Bakry–Emery calculus · Generalized log–Sobolev inequality · Generalized Poincare inequality · Generalized Talagrand inequality · Generalized Yano's formula.

List of symbols

M	Base manifold
$g, (\cdot, \cdot)$	Metric
$\|\cdot\|$	Norm
$\nabla\cdot$	Divergence operator
∇	Gradient operator
Hess	Hessian operator
$\mathcal{P}$	Density manifold
ρ	Probability density
μ	Reference density
$T_\rho\mathcal{P}$	Tangent space

Communicated by Jun Zhang.

✉ Wuchen Li
wuchen@mailbox.sc.edu

1 University of South Carolina, Columbia, SC, United States

$T_\rho^*\mathcal{P}$	Cotangent space
g_ρ	Density manifold metric tensor
$\Delta_h = \nabla \cdot (h\nabla)$	Weighted Laplacian operator
δ	First L^2 variation
δ^2	Second L^2 variation
grad_g	Gradient operator
Hess_g	Hessian operator
$\Gamma_\rho(\cdot, \cdot)$	Christoffel symbol
$(\rho, \sigma) \in T\mathcal{P}$	Tangent bundle
$(\rho, \Phi) \in T^*\mathcal{P}$	Cotangent bundle
$\mathcal{D}_\gamma$	γ-divergence
$\mathcal{I}_\gamma$	γ-Fisher information
$\mathcal{W}_\gamma$	γ-Wasserstein distance
L_γ, L_γ^*	γ-Diffusion process generator
$\Gamma_{\gamma,1}$	γ-Gamma one operator
$\Gamma_{\gamma,2}$	γ-Gamma two operator
κ	Log Sobolev constant
λ	Poincare constant

1 Introduction

Diffusion hypercontractivity plays essential roles in functional inequalities [5, 14] and information theory [1–3, 10, 11]. Moreover, it can be used in estimating convergence rates of Markov chain Monte Carlo algorithms. Among these studies, Bakry–Emery criteria [4] provide sufficient conditions to derive convergence rates of diffusion processes. Recently, optimal transport provides the other calculation methods on this topic [27]. The probability density space is equipped with an infinite-dimensional Riemannian metric, named Wasserstein metric [24]. The density space with the Wasserstein metric is called density manifold [17]. Diffusion hypercontractivity, a.k.a. Bakry–Emery criteria can be derived from Hessian operators of divergence functionals in density manifold [25]. This study has been extended to general ground metric spaces [22, 26]. On the other hand, the relation between local behavior of diffusion hypercontractivity (such as Poincare inequality) and integral formula, known as Yano's formula [28], has been discovered in [18]. Moreover, it shows a connection between the integration formula on the base manifold and (formal) calculus in the density manifold.

In this paper, we study hypercontractivities (convergence properties) for a class of generalized diffusion processes in [21]. Following generalized (mobility) density manifolds proposed in [7, 8, 12, 25] and calculation methods in [19], we derive a one-parameter generalization of Bakry–Emery criteria in Theorem 1. These criteria provide sufficient conditions to derive convergence rates of generalized diffusion processes, which also establishes generalized log–Sobolev and Talagrand inequalities. In addition, a generalized Yano's formula in Theorem 2 is derived, which provides a reference measure-dependent integral formula. We also establish the generalized

Poincare inequality in Corollary 3. In addition, a $PH^{-1}I$ inequality is presented in Theorem 4.

The generalized optimal transport metric spaces have been studied in [12]. They are known as the Dolbeault–Nazaret–Savare space. Many groups have studied associated generalized geodesics [8], and functional inequalities [6, 13]. It is also worth mentioning that the recent preprint [23] discusses entropic regularizations of geodesics on the Dolbeault-Nazaret-Savare space and related gradient flows. Firstly, [13] studies functional inequalities for classical Kolmogorov-Fokker-Planck equations, where the Bakry–Emery criteria are classical. At the same time, they obtain rigorous results in non-smooth settings. In this paper, we build new functional inequalities. First, we introduce new Bakry–Emery criteria, in which the optimal transport type metric in inequalities does not depend on the reference measure. For example, we obtain several functional inequalities related to the classical H^{-1} metric. In addition, [6] formulates divergence-related functional inequalities for a class of drift-diffusion processes. They apply classical Bakry–Emery iterative calculus. In contrast with their results, we introduce a new mean-field type Gamma calculus.

This paper is organized as follows. In Sect. 2, we state the main result of this paper. We establish hypercontractivity for a class of drift-diffusion processes and prove several functional inequalities. A generalized Yano's formula is also derived. In Sect. 3, we formulate the primary tool of the proof. In Sect. 4, we present all proofs. Finally, in Sect. 5, we derive a mean-field type Gamma calculus.

2 Main result

Suppose (M, g) is a compact and smooth finite dimensional Riemannian manifold without boundary. Denote the metric tensor as g, the volume form as dx, the Ricci curvature tensor as Ric, the gradient, divergence, Laplacian operators as ∇, $\nabla\cdot$, Δ, respectively, and the Hessian operator as Hess. For concreteness of the presentation, we assume that $(M, g) = (\mathbb{T}^d, \mathbb{I})$, where $\mathbb{T}^d$ is a d-dimensional torus and $\mathbb{I} \in \mathbb{R}^{d\times d}$ is an identity matrix.

Given a reference probability density function $\mu \in C^\infty(M)$ with $\inf_{x\in M} \mu(x) > 0$, consider the *γ-drift diffusion process*

$$dX_{\gamma,t} = -\frac{\gamma}{\gamma-1}\nabla\mu(X_{\gamma,t})^{\gamma-1}dt + \sqrt{2\mu(X_{\gamma,t})^{\gamma-1}}dB_t,$$

where B_t is the standard Brownian motion in M with the infinitesimal generator

$$L_\gamma\Phi = (\frac{\gamma}{\gamma-1}\nabla\mu^{\gamma-1}, \nabla\Phi) + \mu^{\gamma-1}\Delta\Phi, \quad \Phi \in C^\infty(M).$$

Consider the γ-*divergence functional*[1]

$$\mathcal{D}_\gamma(\rho\|\mu) = \int f(\frac{\rho}{\mu})\mu dx,$$

where $f\colon [0,\infty) \to \mathbb{R}$ satisfies

$$f(\rho) = \begin{cases} \frac{1}{(1-\gamma)(2-\gamma)}(\rho^{2-\gamma}-1) & \gamma \neq 1, \gamma \neq 2; \\ \rho\log\rho & \gamma = 1; \\ -\log\rho & \gamma = 2. \end{cases}$$

Denote the γ-*Fisher information functional*

$$\mathcal{I}_\gamma(\rho\|\mu) = \int \|\nabla\log\frac{\rho}{\mu}\|^2\rho^\gamma\mu^{2\gamma-2}dx.$$

Consider the γ-*Wasserstein distance* proposed in [12][2]

$$\mathcal{W}_\gamma(\rho,\mu) = \inf_\Phi\left\{\int_0^1\sqrt{\int\|\nabla\Phi_t\|^2\rho_t^\gamma dx}dt\colon \partial_t\rho_t + \nabla\cdot(\rho_t^\gamma\nabla\Phi_t) = 0,\ \rho_0 = \rho,\ \rho_1 = \mu\right\},$$

where $\rho_t = \rho(t,x)$, $\Phi_t = \Phi(t,x)$ and the infimum is over all smooth potential function $\Phi \in [0,1]\times M \to \mathbb{R}$.

We next provide sufficient conditions to describe convergence behaviors of γ-drift diffusion processes. We also derive functional inequalities for γ-divergences, γ–Fisher information functionals and γ-Wasserstein distances.

Theorem 1 (Generalized hypercontractivity) *Assume* $\gamma \in [0,1]$. *Suppose there exists a constant* $\kappa > 0$, *such that*

$$\mu^{\gamma-1}Ric - \frac{1}{\gamma-1}Hess\mu^{\gamma-1} - \Delta\mu^{\gamma-1} + \frac{1}{8}\gamma(\gamma-1)\|\nabla\log\mu\|^2\mu^{\gamma-1} \succeq \kappa. \quad (1)$$

Let ρ_0 *be a smooth initial distribution and* ρ_t *be the probability density function of* γ*-drift diffusion process. Then*

$$\mathcal{D}_\gamma(\rho_t\|\mu) \le e^{-2\kappa t}\mathcal{D}_\gamma(\rho_0\|\mu). \quad (2)$$

Moreover, for any smooth probability density function $\rho \in C^\infty(M)$ *with* $\inf_{x\in M}\rho(x) > 0$, *the generalized log–Sobolev inequality holds*

$$\mathcal{D}_\gamma(\rho\|\mu) \le \frac{1}{2\kappa}\mathcal{I}_\gamma(\rho\|\mu). \quad (3)$$

[1] It is often named α-divergence with $\gamma = \frac{3-\alpha}{2}$. We use notation γ for the simplicity of presentation.

[2] When $\gamma = 1$, we remark that the notation of $\mathcal{W}_1$ represents the classical L^2-Wasserstein distance, not the L^1-Wasserstein distance.

In addition, the generalized Talagrand inequality holds

$$\mathcal{W}_\gamma(\rho,\mu) \le \sqrt{\frac{2\mathcal{D}_\gamma(\rho\|\mu)}{\kappa}}. \tag{4}$$

Example 1 *(Kullback–Leibler divergence* [25]) Consider $\gamma = 1$, $f(\rho) = \rho\log\rho$. Functional $\mathcal{D}_1(\rho\|\mu) = \int \rho\log\frac{\rho}{\mu}dx$ forms the classical Kullback–Leibler divergence function (relative entropy), and $\mathcal{I}_1 = \int \|\nabla\log\frac{\rho}{\mu}\|^2\rho dx$ is the classical relative Fisher information. In this case,

$$dX_{1,t} = -\nabla\log\mu(X_{1,t})dt + \sqrt{2}B_t,$$

is the classical Langevin process, and $\mathcal{W}_1$ is the classical L^2-Wasserstein distance. Hence condition (1) forms

$$Ric - Hess\log\mu \succeq \kappa, \quad \kappa > 0,$$

which is the classical Bakry–Emery criterion. Under this condition, the distribution of drift diffusion process $X_{1,t}$ converges to μ. The log–Sobolev inequality (3) holds

$$\int \rho\log\frac{\rho}{\mu}dx \le \frac{1}{2\kappa}\int \|\nabla\log\frac{\rho}{\mu}\|^2\rho dx.$$

The Talagrand inequality holds

$$\mathcal{W}_1(\rho,\mu) \le \sqrt{\frac{2\mathcal{D}_1(\rho\|\mu)}{\kappa}}.$$

If $M = \mathbb{T}^d$ with $\mathrm{Ric} = 0$ and we denote $\mu(x) = e^{-V(x)}$, the condition (1) forms

$$HessV \succeq \kappa\mathbb{I}.$$

Example 2 *(Pearson divergence)* Consider $\gamma = 0$, $f(\rho) = \frac{1}{2}(\rho^2 - 1)$. Here $\mathcal{D}_0 = \frac{1}{2}\int(\frac{\rho}{\mu} - 1)^2\mu dx$ is named Pearson divergence function, $\mathcal{I}_0 = \int \|\nabla\log\frac{\rho}{\mu}\|^2\mu^{-2}dx$ is the 0-Fisher information and

$$dX_{0,t} = \sqrt{2\mu^{-1}(X_t)}dB_t,$$

is the 0–drift diffusion process. The condition (1) forms

$$\mu^{-1}\mathrm{Ric} + \mathrm{Hess}\mu^{-1} - \Delta\mu^{-1} \succeq \kappa, \quad \kappa > 0.$$

Under this condition, the distribution of drift diffusion process $X_{0,t}$ converges to μ. The generalized log–Sobolev inequality (3) holds

$$\frac{1}{2}\int(\frac{\rho}{\mu} - 1)^2\mu dx \le \frac{1}{2\kappa}\int \|\nabla\log\frac{\rho}{\mu}\|^2\mu^{-2}dx.$$

We next show a new integration identity. This follows from the proof of Theorem 1.

Theorem 2 (Generalized Yano's formula) *Denote* $\Phi \in C^\infty(M)$. *Then*

$$\begin{aligned}
&\int \mu^{-1}\Big(\nabla\cdot(\mu^\gamma \nabla\Phi)\Big)^2 dx\\
&=\int \mu^\gamma\Big\{(\mu^{\gamma-1}Ric-\Delta\mu^{\gamma-1}-\frac{1}{\gamma-1}Hess\mu^{\gamma-1})(\nabla\Phi,\nabla\Phi)+\mu^{\gamma-1}\|Hess\Phi\|^2\\
&\quad+\gamma(\gamma-1)\mu^{\gamma-1}\Big((\nabla\log\mu,\nabla\Phi)^2-\|\nabla\log\mu\|^2\|\nabla\Phi\|^2\Big)\Big\}dx.
\end{aligned}$$

Remark 1 If $\mu(x)$ is a uniform measure, i.e. $\mu(x) = 1$, the above formula is the classical Yano's formula [28]:

$$\int (\Delta\Phi)^2 dx = \int \Big\{\mathrm{Ric}(\nabla\Phi,\nabla\Phi)+\|\mathrm{Hess}\Phi\|^2\Big\}dx.$$

Theorem 2 extends these classical Yano's formulas with general volume measure μ and power constant γ. For example, when $\gamma = 0$, we obtain

$$\begin{aligned}
&\int \mu^{-1}(\Delta\Phi)^2 dx\\
&=\int \Big\{(\mu^{-1}\mathrm{Ric}-\Delta\mu^{-1}+\mathrm{Hess}\mu^{-1})(\nabla\Phi,\nabla\Phi)+\mu^{-1}\|\mathrm{Hess}\Phi\|^2\Big\}dx.
\end{aligned}$$

We then derive a Poincare type inequality, which applies the generalized Yano's formula.

Corollary 3 (Generalized Poincare inequality) *Suppose one of the following condition holds. If there exists a constant* $\lambda > 0$*, such that when* $\gamma \in [0, 1]$,

$$\mu^{\gamma-1}Ric-\Delta\mu^{\gamma-1}-\frac{1}{\gamma-1}Hess\mu^{\gamma-1}\succeq\lambda.$$

Or when $\gamma \in [1,\infty)\cup(-\infty,0]$,

$$\mu^{\gamma-1}Ric-\Delta\mu^{\gamma-1}-\frac{1}{\gamma-1}Hess\mu^{\gamma-1}-\gamma(\gamma-1)\|\nabla\log\mu\|^2\mu^{\gamma-1}\succeq\lambda.$$

Then

$$\int f^2\mu dx \le \frac{1}{\lambda}\int \|\nabla f\|^2\mu^\gamma dx, \tag{5}$$

for any $f \in C^\infty(M)$ *with* $\int f\mu dx = 0$.

Remark 2 When $\gamma = 1$, Corollary 3 recovers the classical Poincare inequality

$$\int f^2 \mu dx \leq \frac{1}{\lambda} \int \|\nabla f\|^2 \mu dx, \quad \int f \mu dx = 0.$$

Remark 3 We derive generalized Poincare inequalities, which follow approximations of generalized log–Sobolev inequalities. Denote $\rho = \mu + \epsilon h$, where $h = f\mu$ and $\int h dx = 0$. The L.H.S. of (5) is from the second order expansion (Hessian metric) of γ-divergence $\mathcal{D}_\gamma(\rho\|\mu)$:

$$\mathcal{D}_\gamma(\mu + \epsilon h\|\mu) = \frac{\epsilon^2}{2} \int f^2 \mu dx + o(\epsilon^2).$$

While the R.H.S. of (5) is from the second order approximation in term of ϵ for the γ–relative Fisher information:

$$\mathcal{I}_\gamma(\mu + \epsilon h\|\mu) = \epsilon^2 \int (\nabla f)^2 \mu^\gamma dx + o(\epsilon^2).$$

Example 3 *(Reverse Kullback–Leibler divergence)* Consider $\gamma = 2$, $f(\rho) = -\log \rho$. Note that $\mathcal{D}_2(\rho\|\mu) = -\int \mu \log \frac{\rho}{\mu} dx$ is named reverse Kullback–Leibler divergence or cross entropy. In this case, the condition in Corollary 3 forms

$$\mu \mathrm{Ric} - \Delta\mu - \mathrm{Hess}\mu - 2\|\nabla \log \mu\|^2 \mu \succeq \lambda, \quad \lambda > 0,$$

Under this condition, the generalized Poincare inequality holds

$$\int f^2 \mu dx \leq \frac{1}{\lambda} \int \|\nabla f\|^2 \mu^2 dx,$$

where $f \in C^\infty(M)$ and $\int f \mu dx = 0$. Again, if $M = \mathbb{T}^d$ with $\mathrm{Ric} = 0$, then condition in Corollary 3 forms

$$-\Delta\mu\mathbb{I} - \mathrm{Hess}\mu - 2\|\nabla \log \mu\|^2 \mu \mathbb{I} \succeq \lambda \mathbb{I}.$$

We last note that when $\gamma = 0$, the γ-Wasserstein distance is exactly the H^{-1} distance:

$$\mathcal{W}_0(\rho, \mu) = H^{-1}(\rho, \mu),$$

where H^{-1} is the negative Sobolev distance between ρ and μ, i.e.,

$$H^{-1}(\rho, \mu) = \sqrt{\int \big(\rho - \mu, \Delta^{-1}(\rho - \mu)\big) dx}.$$

We can prove an inequality among Pearson divergence (P), H^{-1} metric and 0-Fisher information (I), namely $PH^{-1}I$ inequality.

Theorem 4 (Inequalities for H^{-1} metric) *Suppose $\mu^{-1}Ric + Hess\mu^{-1} - \Delta\mu^{-1} \succeq \kappa$, where $\kappa \in \mathbb{R}$, then $PH^{-1}I$ inequality holds*

$$D_0(\rho\|\mu) \le \sqrt{\mathcal{I}_0(\rho\|\mu)}H^{-1}(\rho,\mu) - \frac{\kappa}{2}H^{-1}(\rho,\mu)^2.$$

In addition, if $\kappa \ge 0$, then H^{-1}-Talagrand inequality holds

$$H^{-1}(\rho,\mu) \le \sqrt{\frac{2\mathcal{D}_0(\mu\|\nu)}{\kappa}},$$

Remark 4 If $\kappa > 0$, then $PH^{-1}I$ inequality shows

$$\mathcal{D}_0(\rho\|\mu) \le \sqrt{\mathcal{I}_0(\rho\|\mu)}H^{-1}(\rho,\mu).$$

Using the fact that $H^{-1}(\rho,\mu) \le \sqrt{\frac{2\mathcal{D}_0(\mu\|\nu)}{\kappa}}$ and $\mathcal{D}_0(\rho\|\mu) \le \frac{1}{2\kappa}\mathcal{I}_0(\rho\|\mu)$, we have

$$H^{-1}(\rho,\mu) \le \frac{1}{\kappa}\sqrt{\mathcal{I}_0(\rho\|\mu)}.$$

We note that $PH^{-1}I$ inequality is an analog of inequalities among $\mathcal{D}_1$ (H), Wasserstein-2 metric and 1–Fisher information, known as HWI inequality; see details in [25]. In other words, we generalize the HWI inequality into the one on H^{-1} metric space.

3 Generalized Density manifold

In this section, we introduce the tool to prove above results. We first review a class of Riemannian metrics in probability space, introduced by $\gamma-$Wasserstein distance [12]. We then formally present its Riemannian calculus, including gradient and Hessian operators. Using gradient operators in this metric, we last study the convergence behaviors of γ-divergences and γ–Fisher informationals along with γ-drift diffusion processes.

3.1 Density manifold and its Riemannian calculus

Consider the set of smooth and strictly positive densities

$$\mathcal{P} = \Big\{\rho \in C^\infty(M) \colon \rho(x) > 0,\ \int \rho(x)dx = 1\Big\}.$$

Denote the tangent space of $\mathcal{P}$ at $\rho \in \mathcal{P}$ as

$$T_\rho\mathcal{P} = \Big\{\sigma \in C^\infty(M) \colon \int \sigma(x)dx = 0\Big\}.$$

Consider the γ–Wasserstein metric tensor below.

Definition 5 The inner product $g_\rho\colon T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to\mathbb{R}$ is defined as for any σ_1 and $\sigma_2\in T_\rho\mathcal{P}$:

$$g_\rho(\sigma_1,\sigma_2)=\int\Big(\sigma_1,(-\Delta_{\rho^\gamma})^{-1}\sigma_2\Big)dx,$$

where $\Delta_{\rho^\gamma}=\nabla\cdot(\rho^\gamma\nabla)$ is a weighted elliptic operator. In addition, denote Φ_1, $\Phi_2\in C^\infty(M)/\mathbb{R}=T^*_\rho\mathcal{P}$, with $\sigma_i=-\Delta_{\rho^\gamma}\Phi_i$, $i=1,2$. Then

$$g_\rho(\sigma_1,\sigma_2)=\int(\nabla\Phi_1,\nabla\Phi_2)\rho^\gamma dx.$$

An important observation is that if $\gamma=0$, the proposed $\mathcal{W}_0$ metric is the H^{-1} metric [12]. If $\gamma=1$, the proposed $\mathcal{W}_1$ metric is the L^2-Wasserstein metric [17, 24].

We also note that the characterization of geodesics in $(\mathcal{P},g)$ has been studied in [7, 8, 12]. In this paper, we focus on the Riemannian calculus for density manifold $(\mathcal{P},g)$, using both (ρ,σ) in tangent bundle and (ρ,Φ) in cotangent bundle.

Proposition 6 *The Christoffel symbol* $\Gamma_\rho\colon T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to T_\rho\mathcal{P}$ *in* $(\mathcal{P},g)$ *satisfies*

$$\begin{aligned}\Gamma_\rho(\sigma_1,\sigma_2)&=-\frac{\gamma}{2}\Big\{\Delta_{\rho^{\gamma-1}\sigma_1}\Delta^{-1}_{\rho^\gamma}\sigma_2+\Delta_{\rho^{\gamma-1}\sigma_2}\Delta^{-1}_{\rho^\gamma}\sigma_1+\Delta_{\rho^\gamma}\big((\nabla\Delta^{-1}_{\rho^\gamma}\sigma_1,\nabla\Delta^{-1}_{\rho^\gamma}\sigma_2)\rho^{\gamma-1}\big)\Big\}\\&=-\frac{\gamma}{2}\Big\{\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_1}\Phi_2+\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_2}\Phi_1+\Delta_{\rho^\gamma}\big((\nabla\Phi_1,\nabla\Phi_2)\rho^{\gamma-1}\big)\Big\},\end{aligned}$$

where $\sigma_i=-\Delta_{\rho^\gamma}\Phi_i$, $i=1,2$, *and*

$$\Delta_{\rho^{\gamma-1}\sigma_1}\Delta^{-1}_{\rho^\gamma}\sigma_2=\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_1}\Phi_2=\nabla\cdot(\rho^{\gamma-1}\nabla\cdot(\rho^\gamma\nabla\Phi_1)\nabla\Phi_2).$$

Proof The proof follows the study in [19]. We derive the Christoffel symbol by using the Lagrangian formulation of geodesics. Consider the minimization of the geometric action functional in density space:

$$\mathcal{L}(\rho_t,\partial_t\rho_t)=\int_0^1\int\frac{1}{2}(\partial_t\rho_t,(-\Delta_{\rho_t^\gamma})^{-1}\partial_t\rho_t)dxdt,$$

where $\rho_t=\rho(t,x)$ is a density path with fixed initial-terminal time boundary points ρ_0, ρ_1. The geodesics in $(\mathcal{P},g)$ satisfies the Euler-Lagrange equation

$$\frac{\partial}{\partial t}\delta_{\partial_t\rho_t}\mathcal{L}(\rho_t,\partial_t\rho_t)=\delta_{\rho_t}\mathcal{L}(\rho_t,\partial_t\rho_t)+C(t),\tag{6}$$

i.e.

$$\partial_t(-\Delta^{-1}_{\rho_t^\gamma}\partial_t\rho_t)=\delta_\rho\int\frac{1}{2}(\partial_t\rho,(-\Delta_{\rho_t^\gamma})^{-1}\partial_t\rho_t)dx+C(t),$$

where $C(t)$ is a function depending only on t. Using the fact that

$$\partial_t \Delta^{-1}_{\rho_t^\gamma} = -\Delta^{-1}_{\rho_t^\gamma} \cdot \Delta_{\gamma\rho_t^{\gamma-1}\partial_t\rho_t} \cdot \Delta^{-1}_{\rho_t^\gamma},$$

then equation (6) forms

$$-\Delta^{-1}_{\rho_t^\gamma}\partial_{tt}\rho_t + \Delta^{-1}_{\rho_t^\gamma}\Delta_{\gamma\rho_t^{\gamma-1}\partial_t\rho_t}\Delta^{-1}_{\rho_t^\gamma}\partial_t\rho_t = -\frac{1}{2}\|\nabla\Delta^{-1}_{\rho_t^\gamma}\partial_t\rho_t\|^2\gamma\rho_t^{\gamma-1}.$$

Multiplying both sides with $\Delta_{\rho_t^\gamma}$ and comparing with the geodesics equation, we have

$$\partial_{tt}\rho_t + \Gamma_{\rho_t}(\partial_t\rho_t, \partial_t\rho_t) = 0.$$

Hence we derive the Christoffel symbol. Similarly, we can formulate the Christoffel symbol (raised Christoffel symbol) in term of Φ. $\square$

Proposition 7 *The geodesics equation in $(\mathcal{P}, g)$ satisfies*

$$\partial_{tt}\rho_t - \gamma\Delta_{\rho_t^{\gamma-1}\partial_t\rho_t}\Delta^{-1}_{\rho_t^\gamma}\partial_t\rho_t - \frac{\gamma}{2}\Delta_{\rho_t^\gamma}\left(\|\nabla\Delta^{-1}_{\rho_t^\gamma}\partial_t\rho_t\|^2\rho_t^{\gamma-1}\right) = 0.$$

Denote Legendre transform $\Phi_t = (-\Delta_{\rho_t^\gamma})^{-1}\partial_t\rho_t$. Then the co-geodesics equation satisfies

$$\begin{cases} \partial_t\rho_t + \nabla\cdot(\rho_t^\gamma\nabla\Phi_t) = 0, \\ \partial_t\Phi_t + \dfrac{\gamma}{2}\|\nabla\Phi_t\|^2\rho_t^{\gamma-1} = 0. \end{cases} \tag{7}$$

Remark 5 We remark that analytical properties of geodesics equations (7) in Dolbeault–Nazaret–Savare metric spaces have been studied in [8]. We focus on its formal formulation, using which we derive the Hessian operator in generalized optimal transport spaces.

Proof The geodesics equation follows $\partial_{tt}\rho_t + \Gamma_{\rho_t}(\partial_t\rho_t, \partial_t\rho_t) = 0$. We next demonstrate the Hamiltonian formulation of geodesics flow. Consider the Legendre transform in $(\mathcal{P}, g)$:

$$\mathcal{H}(\rho_t, \Phi_t) = \sup_{\Phi_t\in C^\infty(M)} \int \Phi_t\partial_t\rho_t dx - \mathcal{L}(\rho_t, \partial_t\rho_t).$$

Then $\Phi_t = -\Delta^{-1}_{\rho_t^\gamma}\partial_t\rho_t$, and

$$\mathcal{H}(\rho_t, \Phi_t) = \frac{1}{2}\int \Phi_t(-\Delta_{\rho_t^\gamma}\Phi_t)dx = \frac{1}{2}\int \|\nabla\Phi_t\|^2\rho_t^\gamma dx,$$

where the second equality is from the integration by parts formula. Then the co-geodesic flow satisfies

$$\partial_t \rho_t = \delta_{\Phi_t}\mathcal{H}(\rho_t, \Phi_t), \quad \partial_t \Phi_t = -\delta_{\rho_t}\mathcal{H}(\rho_t, \Phi_t),$$

which is the equation pair (7). □

For the completeness of this paper, we present the Lagrangian coordinates of geodesics in generalized density manifolds.

Proposition 8 (Lagrangian coordinates) *Denote $\rho_t = X_t\#\rho^0$, where $X_t : M \to M$ is the diffeomorphism and # is the push-forward operator. Then geodesic equation* (7) *in term of diffeomorphism mapping X_t satisfies*

$$\frac{d^2}{dt^2}X_t + \frac{\gamma-1}{2}\nabla\left\|\frac{d}{dt}X_t\right\|^2 + (\gamma-1)\frac{d}{dt}X_t\nabla\cdot\left(\frac{d}{dt}X_t\right)$$
$$-\frac{(\gamma-2)(\gamma-1)}{2}\nabla\log\rho_t\left\|\frac{d}{dt}X_t\right\|^2 = 0.$$

Here ∇, $\nabla\cdot$ are gradient and divergence operators w.r.t. X_t.

Remark 6 We present three examples of generalized geodesics in Lagrangian coordinates.

(i) If $\gamma = 1$, the $\mathcal{W}_1$ geodesic equation satisfies

$$\frac{d^2}{dt^2}X_t = 0,$$

which is a well-known result in optimal transport.

(ii) If $\gamma = 2$, the $\mathcal{W}_2$ geodesic equation satisfies

$$\frac{d^2}{dt^2}X_t + \frac{1}{2}\nabla\left\|\frac{d}{dt}X_t\right\|^2 + \frac{d}{dt}X_t\nabla\cdot(\frac{d}{dt}X_t) = 0.$$

(iii) If $\gamma = 0$, the $\mathcal{W}_0$, i.e. H^{-1}, geodesic equation satisfies

$$\frac{d^2}{dt^2}X_t - \frac{1}{2}\nabla\left\|\frac{d}{dt}X_t\right\|^2 - \frac{d}{dt}X_t\nabla\cdot(\frac{d}{dt}X_t) - \nabla\log\rho_t\left\|\frac{d}{dt}X_t\right\|^2 = 0.$$

Proof Denote

$$\frac{d}{dt}X_t = \rho^{\gamma-1}(t, X_t)\nabla\Phi(t, X_t),$$

where (ρ_t, Φ_t) satisfies (7). Then

$$\frac{d^2}{dt^2}X_t(t,x) = \left\{\rho^{\gamma-1}\nabla\partial_t\Phi + \rho^{\gamma-1}\nabla\nabla\Phi\frac{d}{dt}X_t + (\gamma-1)\rho^{\gamma-2}\nabla\Phi\frac{d}{dt}\rho\right\}(t, X_t). \tag{8}$$

Notice the fact that

$$
\begin{aligned}
\frac{d}{dt}\rho(t, X_t) =& \Big\{\partial_t \rho + \nabla\rho \cdot \frac{d}{dt}X_t\Big\}(t, X_t)\\
=& \Big\{\partial_t \rho + \rho^{\gamma-1}\nabla\rho \cdot \nabla\Phi\Big\}(t, X_t)\\
=& \Big\{\partial_t \rho + \nabla\cdot(\rho^\gamma\nabla\Phi) - \rho\nabla\cdot(\rho^{\gamma-1}\nabla\Phi)\Big\}(t, X_t)\\
=& -\Big\{\rho\nabla\cdot(\rho^{\gamma-1}\nabla\Phi)\Big\}(t, X_t),
\end{aligned}
$$

where the equality holds since $\partial_t\rho + \nabla\cdot(\rho^\gamma\nabla\Phi) = 0$. Substituting the above formula and $\partial_t\Phi + \frac{\gamma}{2}\|\nabla\Phi\|^2\rho^{\gamma-1} = 0$ into equation (8), we obtain

$$
\begin{aligned}
\frac{d^2}{dt^2}X_t(t, x) =& \Big\{\rho^{\gamma-1}\nabla(-\frac{\gamma}{2}\|\nabla\Phi\|^2\rho^{\gamma-1}) + \rho^{\gamma-1}\nabla\nabla\Phi\rho^{\gamma-1}\nabla\Phi\\
&+ (\gamma-1)\rho^{\gamma-2}(-\rho\nabla\cdot(\rho^{\gamma-1}\nabla\Phi)\rho\nabla\Phi)\Big\}(t, X_t)\\
=& \Big\{-(\gamma-1)\rho^{\gamma-1}\nabla\nabla\Phi\rho^{\gamma-1}\nabla\Phi - \frac{\gamma}{2}\rho^{\gamma-1}\nabla\rho^{\gamma-1}\|\nabla\Phi\|^2\\
&-(\gamma-1)\rho^{\gamma-1}\nabla\Phi\nabla\cdot(\rho^{\gamma-1}\nabla\Phi)\Big\}(t, X_t)\\
=& \Big\{-\frac{\gamma-1}{2}\nabla\|\rho^{\gamma-1}\nabla\Phi\|^2 + (\gamma-1-\frac{\gamma}{2})\rho^{\gamma-1}\nabla\rho^{\gamma-1}\|\nabla\Phi\|^2\\
&-(\gamma-1)\rho^{\gamma-1}\nabla\Phi\nabla\cdot(\rho^{\gamma-1}\nabla\Phi)\Big\}(t, X_t)\\
=& \Big\{-\frac{\gamma-1}{2}\nabla\|\rho^{\gamma-1}\nabla\Phi\|^2 + (\gamma-1-\frac{\gamma}{2})(\gamma-1)\nabla\log\rho^{\gamma-1}\|\rho^{\gamma-1}\nabla\Phi\|^2\\
&-(\gamma-1)\rho^{\gamma-1}\nabla\Phi\nabla\cdot(\rho^{\gamma-1}\nabla\Phi)\Big\}(t, X_t)\\
=& \Big\{-\frac{\gamma-1}{2}\nabla\left\|\frac{dX_t}{dt}\right\|^2 - (\gamma-1)\frac{d}{dt}X_t\nabla\cdot(\frac{d}{dt}X_t)\\
&+\frac{(\gamma-2)(\gamma-1)}{2}\nabla\log\rho\left\|\frac{d}{dt}X_t\right\|^2\Big\}(t, X_t).
\end{aligned}
$$

In above derivations, we use the fact that the second last equality follows $(\gamma-1)\rho^{\gamma-1}\nabla\log\rho = \nabla\rho^{\gamma-1}$ and $\frac{d}{dt}X_t = (\rho^{\gamma-1}\nabla\Phi)(t, X_t)$. □

Proposition 9 *Consider a functional $\mathcal{F}\colon\mathcal{P}\to\mathbb{R}$.*

(i) The Riemannian gradient operator of $\mathcal{F}$ in $(\mathcal{P}, g)$ satisfies

$$
grad_g\mathcal{F}(\rho) = -\nabla\cdot(\rho^\gamma\nabla\delta\mathcal{F}(\rho)),
$$

where δ is the L^2 first variation w.r.t. ρ. And the squared norm of gradient operator satisfies

$$g_\rho(grad_g\mathcal{F}(\rho), grad_g\mathcal{F}(\rho)) = \int \|\nabla\delta\mathcal{F}(\rho)\|^2\rho^\gamma dx.$$

(ii) The Riemannian Hessian operator of $\mathcal{F}$ in $(\mathcal{P}, g)$ satisfies

$$\begin{aligned}
&Hess_g\mathcal{F}(\rho)(\sigma_1,\sigma_2)\\
&=\int\int \nabla_x\nabla_y\delta^2\mathcal{F}(\rho)(x,y)\nabla\Phi_1(x)\nabla\Phi_2(y)\rho(x)^\gamma\rho(y)^\gamma dxdy\\
&\quad+\gamma\int Hess\delta\mathcal{F}(\rho)(x)(\nabla\Phi_1(x),\nabla\Phi_2(x))\rho(x)^{2\gamma-1}dx\\
&\quad+\frac{\gamma(\gamma-1)}{2}\int\Big\{(\nabla\delta\mathcal{F}(\rho)(x),\nabla\Phi_1(x))(\nabla\Phi_2(x),\frac{\nabla\rho(x)}{\rho(x)})\\
&\qquad\qquad+(\nabla\delta\mathcal{F}(\rho)(x),\nabla\Phi_2(x))(\nabla\Phi_1(x),\frac{\nabla\rho(x)}{\rho(x)})\\
&\qquad\qquad-(\nabla\delta\mathcal{F}(\rho)(x),\frac{\nabla\rho(x)}{\rho(x)})(\nabla\Phi_1(x),\nabla\Phi_2(x))\Big\}\rho(x)^{2\gamma-1}dx,
\end{aligned}\tag{9}$$

where $\sigma_i = -\Delta_{\rho^\gamma}\Phi_i$, $i = 1, 2$, and δ^2 is the L^2 second variation operator w.r.t. ρ.

Remark 7 There are several interesting examples for Hessian operators of $\mathcal{F}$ studied in [7], including linear, interaction potential energies, and entropies.

Proof (i) The Riemannian gradient operator satisfies

$$g_\rho(\mathrm{grad}_g\mathcal{F}(\rho),\sigma) = \int \delta\mathcal{F}(\rho)\sigma dx, \quad \text{for any}\sigma\in T_\rho\mathcal{P}.$$

Then

$$\begin{aligned}
\mathrm{grad}_g\mathcal{F}(\rho) &= \Big((-\Delta_{\rho^\gamma})^{-1}\Big)^{-1}\delta\mathcal{F}(\rho) = -\Delta_{\rho^\gamma}\delta\mathcal{F}(\rho)\\
&= -\nabla\cdot(\rho^\gamma\nabla\delta\mathcal{F}(\rho)).
\end{aligned}$$

(ii) The Riemannian Hessian operator satisfies

$$\begin{aligned}
&\mathrm{Hess}_g\mathcal{F}(\rho)(\sigma_1,\sigma_2)\\
&=\int\int\delta^2\mathcal{F}(\rho)(x,y)\sigma_1(x)\sigma_2(y)dxdy-\int\delta\mathcal{F}(\rho)(x)\Gamma_\rho(\sigma_1,\sigma_2)(x)dx\\
&=\int\int\delta^2\mathcal{F}(\rho)(x,y)\nabla_x\cdot(\rho(x)^\gamma\nabla_x\Phi_1(x))\nabla_y\cdot(\rho(y)^\gamma\nabla_y\Phi_2(y))dxdy \quad (H1)\\
&+\frac{\gamma}{2}\int\delta\mathcal{F}(\rho)(x)\Big\{\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_1}\Phi_2+\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_2}\Phi_1
\end{aligned}$$

$$+ \Delta_{\rho^\gamma}\big((\nabla\Phi_1, \nabla\Phi_2)\rho^{\gamma-1}\big)\Big\}dx. \qquad (H2)$$

We next formulate the following two terms (H1), (H2). Notice the fact that

$$\begin{aligned}(H1) &= \int \delta^2\mathcal{F}(\rho)(x,y)\nabla\cdot(\rho(x)^\gamma\nabla\Phi_1(x))\nabla\cdot(\rho(y)^\gamma\nabla_y\Phi_2(y))dxdy\\ &= \int\int -\nabla_x\delta^2\mathcal{F}(\rho)(x,y)\nabla\Phi_1(x)\rho(x)^\gamma dx\nabla_y\cdot(\rho(y)^\gamma\nabla_y\Phi_2(y))dy\\ &= \int\int \nabla_x\nabla_y\delta^2\mathcal{F}(\rho)(x,y)\nabla\Phi_1(x)\nabla\Phi_2(y)\rho(x)^\gamma\rho(y)^\gamma dxdy,\end{aligned}$$

where the second and third equalities are shown from integration by parts with respect to x, y. In addition, we estimate three terms in (H2).

$$\begin{aligned}&\int \delta\mathcal{F}(\rho)\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_1}\Phi_2 dx\\ &= \int \delta\mathcal{F}(\rho)\nabla\cdot(\rho^{\gamma-1}\nabla\cdot(\rho^\gamma\nabla\Phi_1)\nabla\Phi_2)dx\\ &= -\int (\nabla\delta\mathcal{F}(\rho), \nabla\Phi_2)\nabla\cdot(\rho^\gamma\nabla\Phi_1)\rho^{\gamma-1}dx\\ &= \int \Big(\nabla\big((\nabla\delta\mathcal{F}(\rho), \nabla\Phi_2)\rho^{\gamma-1}\big), \nabla\Phi_1\Big)\rho^\gamma dx\\ &= \int \Big\{\mathrm{Hess}\delta\mathcal{F}(\rho)(\nabla\Phi_1, \nabla\Phi_2) + \mathrm{Hess}\Phi_2(\nabla\Phi_1, \nabla\delta\mathcal{F}(\rho))\Big\}\rho^{2\gamma-1}dx\\ &+ \int (\nabla\delta\mathcal{F}(\rho), \nabla\Phi_2)(\nabla\rho^{\gamma-1}, \nabla\Phi_1)\rho^\gamma dx,\end{aligned}$$

where the second and third equalities from the integration by parts with respect to x. Similarly,

$$\begin{aligned}&\int \delta\mathcal{F}(\rho)(x)\Delta_{\rho^{\gamma-1}\Delta_{\rho^\gamma}\Phi_2}\Phi_1 dx\\ &= \int \Big(\nabla\big((\nabla\delta\mathcal{F}(\rho), \nabla\Phi_1)\rho^{\gamma-1}\big), \nabla\Phi_2\Big)\rho^\gamma dx\\ &= \int \Big\{\mathrm{Hess}\delta\mathcal{F}(\rho)(\nabla\Phi_1, \nabla\Phi_2) + \mathrm{Hess}\Phi_1(\nabla\Phi_2, \nabla\delta\mathcal{F}(\rho))\Big\}\rho^{2\gamma-1}dx\\ &+ \int (\nabla\delta\mathcal{F}(\rho), \nabla\Phi_1)(\nabla\rho^{\gamma-1}, \nabla\Phi_2)\rho^\gamma dx.\end{aligned}$$

And

$$\begin{aligned}&\int \delta\mathcal{F}(\rho)\Delta_{\rho^\gamma}\Big((\nabla\Phi_1, \nabla\Phi_2)\rho^{\gamma-1}\Big)dx\\ &= \int \delta\mathcal{F}(\rho)\nabla\cdot(\rho^\gamma\nabla\big((\nabla\Phi_1, \nabla\Phi_2)\rho^{\gamma-1}\big))dx\end{aligned}$$

$$
\begin{aligned}
&= -\int \Big(\nabla\delta\mathcal{F}(\rho), \nabla\big((\nabla\Phi_1, \nabla\Phi_2)\rho^{\gamma-1}\big)\Big)\rho^\gamma dx \\
&= -\int (\nabla\delta\mathcal{F}(\rho), \nabla\rho^{\gamma-1})(\nabla\Phi_1, \nabla\Phi_2)\rho^\gamma dx \\
&- \int \Big\{\mathrm{Hess}\Phi_1(\nabla\delta\mathcal{F}(\rho), \nabla\Phi_2)\rho^{2\gamma-1} + \mathrm{Hess}\Phi_2(\nabla\delta\mathcal{F}(\rho), \nabla\Phi_1)\rho^{2\gamma-1}\Big\}dx.
\end{aligned}
$$

Plugging the above three terms into (H2), we finish the proof. □

3.2 Gradient systems and γ-drift diffusion process

In this subsection, we present the mathematical connection among Riemannian gradient operators in $(\mathcal{P}, g)$, γ-divergence functional, γ-Fisher information and γ-drift diffusion process, see details in [21]. In a word, given a γ-divergence functional, the Kolomogrov forward operator of γ–drift diffusion process is the negative gradient descent direction in $(\mathcal{P}, g)$. Moreover, the squared gradient norm of γ-divergence functional in $(\mathcal{P}, g)$ forms the γ–Fisher information functional.

Lemma 10 *The following statements hold.*

(i)

$$
L_\gamma^*\rho = -grad_g\mathcal{D}_\gamma(\rho\|\mu),
$$

where L_γ^* is the adjoint operator of L_γ in $L^2(\rho)$.

(ii)

$$
\mathcal{I}_\gamma(\rho\|\mu) = g_\rho(grad_g\mathcal{D}_\gamma(\rho\|\mu), grad_g\mathcal{D}_\gamma(\rho\|\mu)).
$$

Proof We first prove (i). On the one hand, the the Kolomogrov forward operator forms

$$
L_\gamma^*\rho = \nabla\cdot\left(\mu^\gamma\nabla\frac{\rho}{\mu}\right).
$$

We need to show

$$
\int L_\gamma\Phi(x)\rho(x)dx = \int \Phi(x)L_\gamma^*\rho(x)dx.
$$

Notice the fact

$$
\begin{aligned}
\int \rho L_\gamma\Phi(x) &= \int \rho\Big\{(\nabla\Phi, \nabla\mu^{\gamma-1}) + \mu^{\gamma-1}\Delta\Phi - \frac{1}{\gamma-1}(\nabla\Phi, \nabla\mu^{\gamma-1})\Big\}dx \\
&= \int \rho\Big\{\nabla\cdot(\mu^{\gamma-1}\nabla\Phi) - (\nabla\Phi, \nabla\mu)\mu^{\gamma-2}\Big\}dx \\
&= \int -(\nabla\Phi, \nabla\rho)\mu^{\gamma-1} - (\nabla\Phi, \nabla\mu)\mu^{\gamma-2}\rho dx
\end{aligned}
$$

$$
\begin{aligned}
&= -\int (\nabla\Phi, \nabla\frac{\rho}{\mu})\mu^{\gamma} dx \\
&= \int \Phi\nabla\cdot(\mu^{\gamma}\nabla\frac{\rho}{\mu})dx \\
&= \int \Phi(x) L^*_{\gamma}\rho(x)dx,
\end{aligned}
$$

where the second equality follows $\nabla\cdot(\mu^{\gamma-1}\nabla\Phi) = (\nabla\mu^{\gamma-1}, \nabla\Phi) + \mu^{\gamma-1}\Delta\Phi$ and the fourth equality applies the fact that $\nabla\frac{\rho}{\mu} = \mu^{-1}\nabla\rho - \mu^{-2}\rho\nabla\mu$. On the other hand, the negative gradient operator of $\mathcal{D}_{\gamma}(\rho\|\mu)$ in $(\mathcal{P}, g)$ satisfies

$$
\begin{aligned}
-\mathrm{grad}_g\mathcal{D}_{\gamma}(\rho\|\mu) &= \nabla\cdot(\rho^{\gamma}\nabla\delta\mathcal{D}_{\gamma}(\rho\|\mu)) \\
&= \nabla\cdot(\rho^{\gamma}\nabla\frac{1}{1-\gamma}\left(\frac{\rho}{\mu}\right)^{1-\gamma}) \\
&= \nabla\cdot(\rho^{\gamma}\left(\frac{\rho}{\mu}\right)^{-\gamma}\nabla\frac{\rho}{\mu}) \\
&= \nabla\cdot(\mu^{\gamma}\nabla\frac{\rho}{\mu}).
\end{aligned}
$$

Comparing the above steps, we finish the proof.

We next prove (ii). Notice the fact that

$$
\begin{aligned}
g_{\rho}(\mathrm{grad}_g\mathcal{D}_{\gamma}(\rho\|\mu), \mathrm{grad}_g\mathcal{D}_{\gamma}(\rho\|\mu)) &= \int \|\nabla\delta\mathcal{D}_{\gamma}(\rho\|\mu)\|^2\rho^{\gamma}dx \\
&= \int \|\nabla\frac{1}{1-\gamma}\left(\frac{\rho}{\mu}\right)^{1-\gamma}\|^2\rho^{\gamma}dx \\
&= \int \|\nabla\log\frac{\rho}{\mu}\|^2\left(\frac{\rho}{\mu}\right)^{2-2\gamma}\rho^{\gamma}dx \\
&= \int \|\nabla\log\frac{\rho}{\mu}\|^2\rho^{2-\gamma}\mu^{2-2\gamma}dx \\
&= \mathcal{I}_{\gamma}(\rho),
\end{aligned}
$$

where the second equality follows $\frac{1}{1-\gamma}\nabla\left(\frac{\rho}{\mu}\right)^{1-\gamma} = \left(\frac{\rho}{\mu}\right)^{-\gamma}\nabla\frac{\rho}{\mu} = \left(\frac{\rho}{\mu}\right)^{1-\gamma}\nabla\log\frac{\rho}{\mu}$. □

We shall apply the above two relations to derive sufficient conditions for proving convergence rates of γ-diffusion processes. We also prove some generalized functional inequalities.

4 Proof

In this section, we present all proofs of the main result in this paper.

4.1 Sketch of proof

Consider the gradient flow of the γ-divergence functional

$$\partial_t \rho_t = -\mathrm{grad}_g \mathcal{D}_\gamma(\rho_t \| \mu) = L_\gamma^* \rho_t,$$

where ρ_t is the probability density function at any time $t > 0$. Then the first time derivative of γ-divergence along the gradient flow satisfies

$$\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu) = -g_\rho(\mathrm{grad}_g \mathcal{D}_\gamma(\rho_t\|\mu), \mathrm{grad}_g \mathcal{D}_\gamma(\rho_t\|\mu)).$$

And the second time derivative of γ-divergence becomes

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{D}_\gamma(\rho_t\|\mu) &= 2\mathrm{Hess}_g \mathcal{D}_\gamma(\rho_t\|\mu)(\partial_t\rho_t, \partial_t\rho_t) \\ &= 2\mathrm{Hess}_g \mathcal{D}_\gamma(\rho_t\|\mu)(\mathrm{grad}_g \mathcal{D}_\gamma(\rho_t\|\mu), \mathrm{grad}_g \mathcal{D}_\gamma(\rho_t\|\mu)).\end{aligned}$$

If we can find the ratio between the first and second derivative, i.e.,

$$\frac{d^2}{dt^2}\mathcal{D}_\gamma(\rho_t\|\mu) \geq -2\kappa \frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu), \tag{10}$$

then we prove Theorem 1. This is true if we integrate (10) on both sides for $[t, \infty)$, then

$$-\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu) \geq 2\kappa \mathcal{D}_\gamma(\rho_t\|\mu). \tag{11}$$

Following Grownwall's inequality, we obtain the convergence result:

$$\mathcal{D}_\gamma(\rho_t\|\mu) \leq e^{-2\kappa t}\mathcal{D}_\gamma(\rho_0\|\mu).$$

Notice the fact that $\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu) = -\mathcal{I}_\gamma(\rho_t\|\mu_t)$. Inequality (11) satisfies

$$\mathcal{I}_\gamma(\rho_t\|\mu) \geq 2\kappa \mathcal{D}_\gamma(\rho_t\|\mu).$$

By choosing $t = 0$ with arbitrary $\rho_0 \in \mathcal{P}$, the log–Sobolev inequality (3) is proven.

From above arguments, the proof boils down to estimate the ratio in (10). In other words, we need to estimate a constant $\kappa > 0$, such that

$$\begin{aligned}&\mathrm{Hess}_g \mathcal{D}_\gamma(\rho\|\mu)(\mathrm{grad}_g \mathcal{D}_\gamma(\rho\|\mu), \mathrm{grad}_g \mathcal{D}_\gamma(\rho\|\mu)) \\ &\geq \kappa g_\rho(\mathrm{grad}_g \mathcal{D}_\gamma(\rho\|\mu), \mathrm{grad}_g \mathcal{D}_\gamma(\rho\|\mu)).\end{aligned}$$

4.2 Hessian operator estimation

We derive Hessian operators of γ-divergences in $(\mathcal{P}, g)$.

Lemma 11 *(Hessian of γ-divergence in $(\mathcal{P}, g)$) Denote $\sigma = -\Delta_{\rho^\gamma}\Phi$. Then*

$$
\begin{aligned}
& Hess_g \mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma) \\
&= \int \rho^\gamma \Big\{ \big(\mu^{\gamma-1} Ric - \Delta\mu^{\gamma-1} - \frac{1}{\gamma-1} Hess\mu^{\gamma-1}\big)(\nabla\Phi,\nabla\Phi) + \mu^{\gamma-1}\|Hess\Phi\|^2 \\
&\qquad + \gamma(\gamma-1)\mu^{\gamma-1}\Big(\Big(\frac{\nabla\rho}{\rho}, \nabla\Phi\Big)\Big(\frac{\nabla\mu}{\mu}, \nabla\Phi\Big) \\
&\quad - \frac{1}{2}\Big(\frac{\nabla\rho}{\rho}, \frac{\nabla\mu}{\mu} + \frac{\nabla\rho}{\rho}\Big)(\nabla\Phi,\nabla\Phi)\Big)\Big\} dx.
\end{aligned}
$$

Remark 8 In fact, there are several interesting examples for Hessian operators of γ-divergence in density manifold. Some of them reformulate the ones derived in [7]. Denote $\sigma = -\Delta_{\rho^\gamma}\Phi$.

(i) If $\gamma = 1$, then

$$
\mathrm{Hess}_g \mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma) = \int \rho \Big\{ (\mathrm{Ric} - \mathrm{Hess}\log\mu)(\nabla\Phi,\nabla\Phi) + \|\mathrm{Hess}\Phi\|^2 \Big\} dx.
$$

(ii) If $\mu = 1$ is a uniform density function, then

$$
\begin{aligned}
\mathrm{Hess}_g \mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma) = \int \rho^\gamma \Big\{ \Big(\mathrm{Ric} - \frac{1}{2}\gamma(\gamma-1) \\
\Big(\frac{\nabla\rho}{\rho}, \frac{\nabla\rho}{\rho}\Big)\Big)(\nabla\Phi,\nabla\Phi) + \|\mathrm{Hess}\Phi\|^2 \Big\} dx.
\end{aligned}
$$

The above formula has been derived in [7].

(iii) If $\gamma = 0$, then

$$
\begin{aligned}
\mathrm{Hess}_g \mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma) = \int \Big\{ \big(\mu^{-1}\mathrm{Ric} - \Delta\mu^{-1} \\
+ \mathrm{Hess}\mu^{-1}\big)(\nabla\Phi,\nabla\Phi) + \mu^{-1}\|\mathrm{Hess}\Phi\|^2 \Big\} dx.
\end{aligned}
$$

Proof From Proposition 9, we can compute the Hessian operator of γ-divergence. The Hessian operator is derived by taking the second order time derivative of $\mathcal{D}_\gamma(\rho\|\mu)$

along with co-geodesics flow (7). Consider the first order time derivative of the γ-divergence:

$$
\begin{aligned}
\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu) &= \int \delta\mathcal{D}_\gamma(\rho_t\|\mu)\partial_t\rho_t dx \\
&= \int \delta\mathcal{D}_\gamma(\rho_t\|\mu)\Big(-\nabla\cdot(\rho_t^\gamma\nabla\Phi_t)\Big)dx \\
&= \int (\nabla\delta\mathcal{D}_\gamma(\rho_t\|\mu), \nabla\Phi_t)\rho_t^\gamma dx.
\end{aligned}
$$

And the second order time derivative of the γ-divergence satisfies

$$
\begin{aligned}
&\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma) \\
&\quad= \frac{d^2}{dt^2}\mathcal{D}_\gamma(\rho_t\|\mu)|_{t=0} \\
&= \frac{d}{dt}\Big(\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu)\Big)|_{t=0} \\
&= \int (\nabla\frac{d}{dt}\delta\mathcal{D}_\gamma(\rho_t\|\mu), \nabla\Phi_t)\rho_t^\gamma dx|_{t=0} && (A) \\
&+ \int (\nabla\delta\mathcal{D}_\gamma(\rho_t\|\mu), \nabla\partial_t\Phi_t)\rho_t^\gamma dx|_{t=0} && (B) \\
&+ \int (\nabla\delta\mathcal{D}_\gamma(\rho_t\|\mu), \nabla\Phi_t)(\gamma\rho_t^{\gamma-1})\partial_t\rho_t dx|_{t=0} && (C)
\end{aligned}
$$

We estimate (A), (B) and (C) separately. For (A), we denote $\delta^2\mathcal{D}(\rho) = f''(\frac{\rho}{\mu})\frac{1}{\mu} = \rho^{-\gamma}\mu^{\gamma-1}$. Then

$$
\begin{aligned}
(A) &= \int \delta^2\mathcal{D}_\gamma(\rho\|\mu)\Big(\nabla\cdot(\rho^\gamma\nabla\Phi)\Big)^2 dx \\
&= \int \delta^2\mathcal{D}_\gamma(\rho\|\mu)\Big((\nabla\rho^\gamma, \nabla\Phi) + \rho^\gamma\Delta\Phi\Big)^2 dx \\
&= \int \delta^2\mathcal{D}_\gamma(\rho\|\mu)\Big((\nabla\rho^\gamma, \nabla\Phi)^2 + 2(\nabla\rho^\gamma, \nabla\Phi)\rho^\gamma\Delta\Phi + \rho^{2\gamma}(\Delta\Phi)^2\Big) \\
&= \int \rho^{-\gamma}\mu^{\gamma-1}\Big((\nabla\rho^\gamma, \nabla\Phi)^2 + 2(\nabla\rho^\gamma, \nabla\Phi)\rho^\gamma\Delta\Phi + \rho^{2\gamma}(\Delta\Phi)^2\Big)dx \\
&= \int \rho^{-\gamma}\mu^{\gamma-1}\gamma^2\rho^{2\gamma-2}(\nabla\rho, \nabla\Phi)^2 dx + \int 2\rho^{-\gamma}\mu^{\gamma-1}(\nabla\rho^\gamma, \nabla\Phi)\rho^\gamma\Delta\Phi dx \\
&+ \int \mu^{\gamma-1}\rho^\gamma(\Delta\Phi)^2 dx.
\end{aligned}
$$

Hence

$$
\begin{aligned}
(A) &= \int \mu^{\gamma-1}\gamma^2\rho^{\gamma-2}(\nabla\rho, \nabla\Phi)^2 dx + \int 2\mu^{\gamma-1}(\nabla\rho^{\gamma}, \nabla\Phi)\Delta\Phi dx \\
&\quad + \int \mu^{\gamma-1}\rho^{\gamma}(\Delta\Phi)^2 dx \\
&= \gamma^2 \int \mu^{\gamma-1}\rho^{\gamma}\left(\frac{1}{\rho}\nabla\rho, \nabla\Phi\right)^2 dx - 2\int \rho^{\gamma}\nabla\cdot(\mu^{\gamma-1}\nabla\Phi\Delta\Phi)dx \\
&\quad + \int \mu^{\gamma-1}\rho^{\gamma}(\Delta\Phi)^2 dx.
\end{aligned}
$$

We next estimate (B).

$$
\begin{aligned}
(B) &= \int (\nabla\delta\mathcal{D}_{\gamma}(\rho_t\|\mu), \nabla\partial_t\Phi_t)\rho_t^{\gamma} dx|_{t=0} \\
&= \int \partial_t\Phi_t\Big(-\nabla\cdot(\rho_t^{\gamma}\nabla\delta\mathcal{D}_{\gamma}(\rho_t\|\mu))\Big)dx|_{t=0} \\
&= \int \frac{1}{2}(\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}\nabla\cdot(\rho^{\gamma}\nabla\delta\mathcal{D}_{\gamma}(\rho\|\mu))dx \\
&= \frac{1}{2}\int (\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}\nabla\cdot(\rho^{\gamma}\nabla\frac{1}{1-\gamma}(\frac{\rho}{\mu})^{1-\gamma})dx \\
&= \frac{1}{2}\int (\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}\nabla\cdot(\rho^{\gamma}(\frac{\rho}{\mu})^{-\gamma}\nabla\frac{\rho}{\mu})dx \\
&= \frac{1}{2}\int (\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}\nabla\cdot(\mu^{\gamma}\frac{\nabla\rho\mu-\rho\nabla\mu}{\mu^2})dx \\
&= \frac{1}{2}\int (\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}\nabla\cdot(\mu^{\gamma-1}\nabla\rho-\rho\mu^{\gamma-2}\nabla\mu)dx \\
&= -\frac{1}{2}\int \nabla\Big((\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}\Big)\Big(\mu^{\gamma-1}\nabla\rho-\frac{1}{\gamma-1}\rho\nabla\mu^{\gamma-1}\Big)dx \\
&= -\frac{1}{2}\int \Big(\nabla(\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma-1}+(\nabla\Phi, \nabla\Phi)\gamma\nabla\rho^{\gamma-1}\Big) \\
&\quad \Big(\mu^{\gamma-1}\nabla\rho-\frac{1}{\gamma-1}\rho\nabla\mu^{\gamma-1}\Big)dx \\
&= -\frac{1}{2}\int (\nabla(\nabla\Phi, \nabla\Phi), \nabla\rho)\gamma\rho^{\gamma-1}\mu^{\gamma-1}dx \qquad (B1) \\
&\quad -\frac{1}{2}\int (\nabla\rho^{\gamma-1}, \nabla\rho)(\nabla\Phi, \nabla\Phi)\gamma\mu^{\gamma-1}dx \qquad (B2) \\
&\quad +\frac{1}{2(\gamma-1)}\int \Big(\nabla(\nabla\Phi, \nabla\Phi)\gamma\rho^{\gamma}, \nabla\mu^{\gamma-1}\Big)dx \qquad (B3) \\
&\quad +\frac{1}{2(\gamma-1)}\int (\nabla\Phi, \nabla\Phi)\gamma(\nabla\rho^{\gamma-1}\rho, \nabla\mu^{\gamma-1})dx. \qquad (B4)
\end{aligned}
$$

We next derive (B1), (B2), (B3), (B4). Notice the fact that

$$
\begin{aligned}
(B1) &= -\frac{1}{2}\int (\nabla(\nabla\Phi, \nabla\Phi), \nabla\rho)\gamma\rho^{\gamma-1}\mu^{\gamma-1}dx \\
&= -\frac{1}{2}\int \Big(\nabla(\nabla\Phi, \nabla\Phi)\mu^{\gamma-1}, \nabla\rho^{\gamma}\Big)dx \\
&= \frac{1}{2}\int \nabla\cdot\Big(\mu^{\gamma-1}\nabla(\nabla\Phi, \nabla\Phi)\Big)\rho^{\gamma}dx \\
&= \frac{1}{2}\int \Big\{(\nabla\mu^{\gamma-1}, \nabla(\nabla\Phi, \nabla\Phi)) + \mu^{\gamma-1}\Delta(\nabla\Phi, \nabla\Phi)\Big\}\rho^{\gamma}dx,
\end{aligned}
$$

and

$$
\begin{aligned}
(B2) &= -\frac{1}{2}\int (\nabla\Phi, \nabla\Phi)(\gamma\nabla\rho^{\gamma-1}, \mu^{\gamma-1}\nabla\rho)dx \\
&= -\frac{1}{2}\gamma(\gamma-1)\int (\nabla\Phi, \nabla\Phi)\rho^{\gamma-2}\mu^{\gamma-1}(\nabla\rho, \nabla\rho)dx \\
&= -\frac{1}{2}\gamma(\gamma-1)\int \rho^{\gamma}\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)(\frac{1}{\rho}\nabla\rho, \frac{1}{\rho}\nabla\rho)dx.
\end{aligned}
$$

In addition,

$$
(B3) = \frac{\gamma}{2(\gamma-1)}\int \rho^{\gamma}\Big(\nabla(\nabla\Phi, \nabla\Phi), \nabla\mu^{\gamma-1}\Big)dx,
$$

and

$$
\begin{aligned}
(B4) &= \frac{1}{2(\gamma-1)}\int (\nabla\Phi, \nabla\Phi)\gamma(\nabla\rho^{\gamma-1}\rho, \nabla\mu^{\gamma-1})dx \\
&= \frac{1}{2(\gamma-1)}\int (\nabla\Phi, \nabla\Phi)(\gamma(\gamma-1)\rho^{\gamma-2}\rho\nabla\rho, \nabla\mu^{\gamma-1})dx \\
&= \frac{1}{2}\int (\nabla\Phi, \nabla\Phi)(\nabla\rho^{\gamma}, \nabla\mu^{\gamma-1})dx \\
&= -\frac{1}{2}\int \rho^{\gamma}\nabla\cdot\Big((\nabla\Phi, \nabla\Phi)\nabla\mu^{\gamma-1}\Big)dx \\
&= -\frac{1}{2}\int \rho^{\gamma}\Big\{(\nabla(\nabla\Phi, \nabla\Phi), \nabla\mu^{\gamma-1}) + (\nabla\Phi, \nabla\Phi)\Delta\mu^{\gamma-1}\Big\}dx.
\end{aligned}
$$

We last derive (C). Consider

$$
\begin{aligned}
(C) &= \int (\nabla\delta\mathcal{D}_\gamma(\rho_t\|\mu), \nabla\Phi_t)(\gamma\rho_t^{\gamma-1})\partial_t\rho_t dx|_{t=0} \\
&= \int (\nabla\frac{1}{1-\gamma}(\frac{\rho}{\mu})^{1-\gamma}, \nabla\Phi)\gamma\rho^{\gamma-1}\Big(-\nabla\cdot(\rho^\gamma\nabla\Phi)\Big)dx \\
&= \int (\frac{\rho}{\mu})^{-\gamma}(\nabla\frac{\rho}{\mu}, \nabla\Phi)\gamma\rho^{\gamma-1}\Big(-\nabla\cdot(\rho^\gamma\nabla\Phi)\Big)dx \\
&= \gamma\int (\nabla\frac{\rho}{\mu}, \nabla\Phi)\mu^\gamma\Big(-\frac{1}{\rho}\nabla\cdot(\rho^\gamma\nabla\Phi)\Big)dx \\
&= \gamma\int (\frac{\nabla\rho}{\mu} - \frac{\rho\nabla\mu}{\mu^2}, \nabla\Phi)\mu^\gamma\Big(-\frac{1}{\rho}\nabla\cdot(\rho^\gamma\nabla\Phi)\Big)dx \\
&= \gamma\int (\nabla\rho\mu^{\gamma-1} - \mu^{\gamma-2}\rho\nabla\mu, \nabla\Phi)\Big(-\frac{1}{\rho}\nabla\cdot(\rho^\gamma\nabla\Phi)\Big)dx \\
&= -\gamma\int (\nabla\rho\mu^{\gamma-1} - \mu^{\gamma-2}\rho\nabla\mu, \nabla\Phi)\Big((\frac{1}{\rho}\nabla\rho^\gamma, \nabla\Phi) + \rho^{\gamma-1}\Delta\Phi\Big)dx \\
&= -\gamma\int (\nabla\rho\mu^{\gamma-1}, \nabla\Phi)(\frac{1}{\rho}\nabla\rho^\gamma, \nabla\Phi)dx \qquad (C1) \\
&- \gamma\int (\nabla\rho, \nabla\Phi)\rho^{\gamma-1}\mu^{\gamma-1}\Delta\Phi dx \qquad (C2) \\
&+ \gamma\int (\mu^{\gamma-2}\nabla\mu, \nabla\Phi)(\nabla\rho^\gamma, \nabla\Phi)dx \qquad (C3) \\
&+ \gamma\int (\mu^{\gamma-2}\nabla\mu, \nabla\Phi)\rho^\gamma\Delta\Phi dx. \qquad (C4)
\end{aligned}
$$

We estimate (C1), (C2), (C3), (C4) explicitly. Notice the fact that

$$
\begin{aligned}
(C1) &= -\gamma\int (\nabla\rho\mu^{\gamma-1}, \nabla\Phi)(\frac{1}{\rho}\nabla\rho^\gamma, \nabla\Phi)dx \\
&= -\gamma^2\int \mu^{\gamma-1}(\nabla\rho, \nabla\Phi)^2\rho^{\gamma-2}dx \\
&= -\gamma^2\int \mu^{\gamma-1}(\frac{1}{\rho}\nabla\rho, \nabla\Phi)^2\rho^\gamma dx,
\end{aligned}
$$

and

$$
\begin{aligned}
(C2) &= -\gamma\int (\nabla\rho, \nabla\Phi)\rho^{\gamma-1}\mu^{\gamma-1}\Delta\Phi dx \\
&= -\int (\nabla\rho^\gamma, \nabla\Phi)\mu^{\gamma-1}\Delta\Phi dx \\
&= \int \rho^\gamma\nabla\cdot(\mu^{\gamma-1}\Delta\Phi\nabla\Phi)dx.
\end{aligned}
$$

In addition,

$$
\begin{aligned}
(C3)+(C4) &= \gamma \int (\mu^{\gamma-2}\nabla\mu, \nabla\Phi)\nabla\cdot(\rho^{\gamma}\nabla\Phi)dx \\
&= -\frac{\gamma}{\gamma-1}\int \rho^{\gamma}\Big(\nabla(\nabla\mu^{\gamma-1},\nabla\Phi),\nabla\Phi\Big)dx.
\end{aligned}
$$

We now summarize all above formulas below.

$$
\begin{aligned}
&\mathrm{Hess}_g\mathcal{D}_{\gamma}(\rho\|\mu)(\sigma,\sigma) \\
=&\frac{d^2}{dt^2}\mathcal{D}_{\gamma}(\rho_t\|\mu)|_{t=0} \\
&= (A)+(B)+(C) \\
&= (A)+(B1)+(B2)+(B3)+(B4)+(C1)+(C2)+(C3)+(C4) \\
&= \gamma^2\int \mu^{\gamma-1}\rho^{\gamma}\left(\frac{1}{\rho}\nabla\rho,\nabla\Phi\right)^2dx \\
&-2\int \rho^{\gamma}\nabla\cdot(\mu^{\gamma-1}\nabla\Phi\Delta\Phi)dx \\
&+\int \mu^{\gamma-1}\rho^{\gamma}(\Delta\Phi)^2dx \\
&+\frac{1}{2}\int\Big\{(\nabla\mu^{\gamma-1},\nabla(\nabla\Phi,\nabla\Phi))+\mu^{\gamma-1}\Delta(\nabla\Phi,\nabla\Phi)\Big\}\rho^{\gamma}dx \\
&-\frac{1}{2}\gamma(\gamma-1)\int\rho^{\gamma}\mu^{\gamma-1}(\nabla\Phi,\nabla\Phi)(\frac{1}{\rho}\nabla\rho,\frac{1}{\rho}\nabla\rho)dx \\
&+\frac{\gamma}{2(\gamma-1)}\int\rho^{\gamma}\Big(\nabla(\nabla\Phi,\nabla\Phi),\nabla\mu^{\gamma-1}\Big)dx \\
&-\frac{1}{2}\int\rho^{\gamma}\Big\{(\nabla(\nabla\Phi,\nabla\Phi),\nabla\mu^{\gamma-1})+(\nabla\Phi,\nabla\Phi)\Delta\mu^{\gamma-1}\Big\}dx \\
&-\gamma^2\int\mu^{\gamma-1}(\frac{1}{\rho}\nabla\rho,\nabla\Phi)^2\rho^{\gamma}dx \\
&+\int\rho^{\gamma}\nabla\cdot(\mu^{\gamma-1}\Delta\Phi\nabla\Phi)dx \\
&-\frac{\gamma}{\gamma-1}\int\rho^{\gamma}\Big(\nabla(\nabla\mu^{\gamma-1},\nabla\Phi),\nabla\Phi\Big)dx \\
&=\int\rho^{\gamma}\Big\{-\nabla\cdot(\mu^{\gamma-1}\nabla\Phi\Delta\Phi)+\mu^{\gamma-1}(\Delta\Phi)^2+\frac{1}{2}\mu^{\gamma-1}\Delta(\nabla\Phi,\nabla\Phi) \qquad (D) \\
&\quad -\frac{1}{2}\Delta\mu^{\gamma-1}(\nabla\Phi,\nabla\Phi)-\frac{\gamma}{\gamma-1}\mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi,\nabla\Phi) \\
&\quad -\frac{1}{2}\gamma(\gamma-1)(\frac{\nabla\rho}{\rho},\frac{\nabla\rho}{\rho})(\nabla\Phi,\nabla\Phi)\mu^{\gamma-1}\Big\}dx.
\end{aligned}
\tag{12}
$$

Note the fact that

$$\begin{aligned}
(D) &= -\nabla\cdot(\mu^{\gamma-1}\nabla\Phi\Delta\Phi) + \mu^{\gamma-1}(\Delta\Phi)^2 + \frac{1}{2}\mu^{\gamma-1}\Delta(\nabla\Phi,\nabla\Phi)\\
&= -(\nabla\mu^{\gamma-1},\nabla\Phi)\Delta\Phi - \mu^{\gamma-1}(\Delta\Phi)^2 - \mu^{\gamma-1}(\nabla\Delta\Phi,\nabla\Phi)\\
&+ \mu^{\gamma-1}(\Delta\Phi)^2 + \frac{1}{2}\mu^{\gamma-1}\Delta(\nabla\Phi,\nabla\Phi)\\
&= -(\nabla\mu^{\gamma-1},\nabla\Phi)\Delta\Phi + \mu^{\gamma-1}\Big\{\frac{1}{2}\Delta(\nabla\Phi,\nabla\Phi) - (\nabla\Phi,\nabla\Delta\Phi)\Big\}\\
&= -(\nabla\mu^{\gamma-1},\nabla\Phi)\Delta\Phi + \mu^{\gamma-1}\Big\{\mathrm{Ric}(\nabla\Phi,\nabla\Phi) + \|\mathrm{Hess}\Phi\|^2\Big\},
\end{aligned}$$

where the last equality is from Bochner's formula, i.e.,

$$\frac{1}{2}\Delta(\nabla\Phi,\nabla\Phi) - (\nabla\Phi,\nabla\Delta\Phi) = \mathrm{Ric}(\nabla\Phi,\nabla\Phi) + \|\mathrm{Hess}\Phi\|^2.$$

By substituting (D) into (12), we obtain

$$\begin{aligned}
&\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(V_\Phi,V_\Phi)\\
&= \int \rho^\gamma\mu^{\gamma-1}\Big\{(\mathrm{Ric}(\nabla\Phi,\nabla\Phi) + \|\mathrm{Hess}\Phi\|^2\Big\}dx\\
&+ \frac{1}{2}\int \rho^\gamma\Delta\mu^{\gamma-1}(\nabla\Phi,\nabla\Phi)dx\\
&- \frac{r}{\gamma-1}\int \rho^\gamma\mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi,\nabla\Phi)dx\\
&- \frac{1}{2}\gamma(\gamma-1)\int \rho^\gamma\mu^{\gamma-1}(\frac{\nabla\rho}{\rho},\frac{\nabla\rho}{\rho})(\nabla\Phi,\nabla\Phi)dx\\
&- \int \rho^\gamma(\nabla\mu^{\gamma-1},\nabla\Phi)\Delta\Phi dx. \qquad (E)
\end{aligned} \tag{13}$$

We lastly reformulate the term (E):

$$\begin{aligned}
(E) &= -\int \rho^\gamma(\nabla\mu^{\gamma-1},\nabla\Phi)\nabla\cdot(\nabla\Phi)dx\\
&= \int \Big(\nabla\big(\rho^\gamma(\nabla\mu^{\gamma-1},\nabla\Phi)\big),\nabla\Phi\Big)dx\\
&= \int (\nabla\rho^\gamma,\nabla\Phi)(\nabla\mu^{\gamma-1},\nabla\Phi)dx + \int \rho^\gamma(\nabla(\nabla\mu^{\gamma-1},\nabla\Phi),\nabla\Phi)dx\\
&= \int (\nabla\rho^\gamma,\nabla\Phi)(\nabla\mu^{\gamma-1},\nabla\Phi)dx + \int \rho^\gamma\mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi,\nabla\Phi)dx\\
&+ \int \rho^\gamma\mathrm{Hess}\Phi(\nabla\Phi,\nabla\mu^{\gamma-1})dx. \qquad (E1)
\end{aligned}$$

Notice that (E1) has a formulation:

$$
\begin{aligned}
(E1) &= \int \rho^\gamma \mathrm{Hess}\Phi(\nabla\Phi, \nabla\mu^{\gamma-1})dx \\
&= \int \rho^\gamma \Big(\nabla\big(\frac{1}{2}(\nabla\Phi)^2\big), \nabla\mu^{\gamma-1}\Big)dx \\
&= -\frac{1}{2}\int \nabla\cdot(\rho^\gamma\nabla\mu^{\gamma-1})(\nabla\Phi)^2 dx \\
&= -\frac{1}{2}\int \Big((\nabla\rho^\gamma, \nabla\mu^{\gamma-1}) + \rho^\gamma\Delta\mu^{\gamma-1}\Big)(\nabla\Phi)^2 dx,
\end{aligned}
$$

where the second equality holds by the fact that $\mathrm{Hess}\Phi\nabla\Phi = \nabla\nabla\Phi\nabla\Phi = \frac{1}{2}\nabla(\nabla\Phi)^2$. Substituting (E1) into (E), we obtain

$$
\begin{aligned}
(E) = &\int (\nabla\rho^\gamma, \nabla\Phi)(\nabla\mu^{\gamma-1}, \nabla\Phi)dx - \frac{1}{2}\int (\nabla\rho^\gamma, \nabla\mu^{\gamma-1})(\nabla\Phi, \nabla\Phi)dx \\
&+ \int \rho^\gamma \mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx - \frac{1}{2}\int \rho^\gamma\Delta\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx \\
&= \gamma(\gamma-1)\int \rho^\gamma\mu^{\gamma-1}\Big\{(\frac{\nabla\rho}{\rho}, \nabla\Phi)(\frac{\nabla\mu}{\mu}, \nabla\Phi) - \frac{1}{2}(\frac{\nabla\rho}{\rho}, \frac{\nabla\mu}{\mu})(\nabla\Phi, \nabla\Phi)\Big\}dx \\
&+ \int \rho^\gamma \mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx - \frac{1}{2}\int \rho^\gamma\Delta\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx.
\end{aligned}
$$

Plugging the above formula into (13), we derive

$$
\begin{aligned}
&\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(V_\Phi, V_\Phi) \\
&= \int \rho^\gamma\mu^{\gamma-1}\Big\{\mathrm{Ric}(\nabla\Phi, \nabla\Phi) + \|\mathrm{Hess}\Phi\|^2\Big\}dx - \frac{1}{2}\int \rho^\gamma\Delta\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx \\
&- \frac{r}{\gamma-1}\int \rho^\gamma \mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx \\
&- \frac{1}{2}\gamma(\gamma-1)\int \rho^\gamma\mu^{\gamma-1}(\frac{\nabla\rho}{\rho}, \frac{\nabla\rho}{\rho})(\nabla\Phi, \nabla\Phi)dx \\
&+ \gamma(\gamma-1)\int \rho^\gamma\mu^{\gamma-1}\Big\{(\frac{\nabla\rho}{\rho}, \nabla\Phi)(\frac{\nabla\mu}{\mu}, \nabla\Phi) - \frac{1}{2}(\frac{\nabla\rho}{\rho}, \frac{\nabla\mu}{\mu})(\nabla\Phi, \nabla\Phi)\Big\}dx \\
&+ \int \rho^\gamma \mathrm{Hess}\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx - \frac{1}{2}\int \rho^\gamma\Delta\mu^{\gamma-1}(\nabla\Phi, \nabla\Phi)dx \\
&= \int \rho^\gamma\Big\{\big(\mu^{\gamma-1}\mathrm{Ric} - \Delta\mu^{\gamma-1} - \frac{1}{\gamma-1}\mathrm{Hess}\mu^{\gamma-1}\big)(\nabla\Phi, \nabla\Phi) + \mu^{\gamma-1}\|\mathrm{Hess}\Phi\|^2 \\
&\quad + \gamma(\gamma-1)\mu^{\gamma-1}\Big((\frac{\nabla\rho}{\rho}, \nabla\Phi)(\frac{\nabla\mu}{\mu}, \nabla\Phi) - \frac{1}{2}(\frac{\nabla\rho}{\rho}, \frac{\nabla\mu}{\mu} + \frac{\nabla\rho}{\rho})(\nabla\Phi, \nabla\Phi)\Big)\Big\}dx.
\end{aligned}
$$

□

We observe that the Hessian operator in $(\mathcal{P}, g)$ contains more terms than the one in classical L^2-Wasserstein space. In the classical case, i.e., $\gamma = 1$, there is no interaction bilinear functional between the Hessian operator and the squared gradient norm. In this paper, we overcome this by the following estimations.

Denote the bilinear form below.

$$J(\Phi,\Phi)=(\frac{\nabla\rho}{\rho},\nabla\Phi)(\frac{\nabla\mu}{\mu},\nabla\Phi)-\frac{1}{2}(\frac{\nabla\rho}{\rho},\frac{\nabla\rho}{\rho}+\frac{\nabla\mu}{\mu})(\nabla\Phi,\nabla\Phi).$$

Lemma 12 *Denote $\delta\mathcal{D}_\gamma(\rho\|\mu)=\frac{1}{1-\gamma}(\frac{\rho}{\mu})^{1-\gamma}$. Then for any $\rho\in\mathcal{P}$,*

$$\frac{J(\delta\mathcal{D}_\gamma(\rho\|\mu),\delta\mathcal{D}_\gamma(\rho\|\mu))}{(\nabla\delta\mathcal{D}_\gamma(\rho\|\mu),\nabla\delta\mathcal{D}_\gamma(\rho\|\mu))}\in(-\infty,\frac{1}{8}\|\nabla\log\mu\|^2].$$

Proof The proof is based on an estimation for the bilinear form J. Note that

$$\nabla\delta\mathcal{D}_\gamma(\rho\|\mu)=(\frac{\rho}{\mu})^{-\gamma}\nabla\frac{\rho}{\mu}=(\frac{\rho}{\mu})^{1-\gamma}\nabla\log\frac{\rho}{\mu}.$$

Then

$$\begin{aligned}J_1&:=J(\delta\mathcal{D}_\gamma(\rho\|\mu),\delta\mathcal{D}_\gamma(\rho\|\mu))\\&=\Big\{(\nabla\log\rho,\nabla\log\frac{\rho}{\mu})(\nabla\log\mu,\nabla\log\frac{\rho}{\mu})\\&-\frac{1}{2}(\nabla\log\rho,\nabla\log\rho+\nabla\log\mu)(\nabla\log\frac{\rho}{\mu},\nabla\log\frac{\rho}{\mu})\Big\}(\frac{\rho}{\mu})^{2-2\gamma},\end{aligned}$$

and

$$J_2:=(\nabla\delta\mathcal{D}_\gamma(\rho\|\mu),\nabla\delta\mathcal{D}_\gamma(\rho\|\mu))=(\nabla\log\frac{\rho}{\mu},\nabla\log\frac{\rho}{\mu})(\frac{\rho}{\mu})^{2-2\gamma}.$$

Denote $\nabla\log\frac{\rho}{\mu}=a$, $\nabla\log\mu=a_0$. Then $\nabla\log\rho=a+a_0$. Thus

$$\begin{aligned}\frac{J_1}{J_2}&=\frac{(a+a_0,a)(a_0,a)-\frac{1}{2}(a+a_0,a+2a_0)(a,a)}{(a,a)}\\&=\frac{(a,a)(a_0,a)+(a_0,a)^2-\frac{1}{2}[(a,a)+3(a_0,a)+2(a_0,a_0)](a,a)}{(a,a)}\\&=\frac{(a,a)(a_0,a)+(a_0,a)^2-\frac{1}{2}(a,a)^2-\frac{3}{2}(a_0,a)(a,a)-(a_0,a_0)(a,a)}{(a,a)}\\&=\frac{(a_0,a)^2-\frac{1}{2}(a,a)^2-\frac{1}{2}(a_0,a)(a,a)-(a_0,a_0)(a,a)}{(a,a)}.\end{aligned}$$

We further denote $\cos\theta = \frac{(a_0,a)}{\|a\|\|a_0\|}$. Then

$$\begin{aligned}\frac{J_1}{J_2} &= \frac{\|a\|^2\|a_0\|^2\cos^2\theta - \frac{1}{2}\|a\|^4 - \frac{1}{2}\|a_0\|\|a\|^3\cos\theta - \|a_0\|^2\|a\|^2}{\|a\|^2}\\ &= \|a_0\|^2(\cos^2\theta - 1) - \frac{1}{2}\|a\|^2 - \frac{1}{2}\|a\|\|a_0\|\cos\theta\\ &= \|a_0\|^2(\frac{9}{8}\cos^2\theta - 1) - \frac{1}{2}\big(\|a\| + \frac{1}{2}\|a_0\|\cos\theta\big)^2\\ &\le \frac{1}{8}\|a_0\|^2,\end{aligned}$$

which finishes the proof. □

4.3 Proof

Proof of Theorem 1 Firstly, following Lemma 11 and Lemma 12, we prove that condition (1) implies both the convergence result (10) and the functional inequality (3).

Secondly, the generalized Talagrand inequality (4) follows directly from the gradient flow interpolation of inequality in Proposition 1 of [25]. For the completeness of this paper, we still present it here. Consider the real value function

$$\Psi(t) = \mathcal{W}_\gamma(\rho_0, \rho_t) + \sqrt{\frac{2\mathcal{D}_\gamma(\rho_t\|\mu)}{\kappa}},$$

where $\rho_t = \rho(t,\cdot)$ is the density function at time t. Notice that $\Psi(0) = \mathcal{W}(\rho_0,\mu)$ and $\lim_{t\to\infty}\Psi(t) = \sqrt{\frac{2\mathcal{D}_\gamma(\rho_t\|\mu)}{\kappa}}$, since $\mathcal{D}_\gamma(\rho_t\|\mu) \to 0$ following (10).

We next claim $\frac{d}{dt}\Psi(t) \le 0$. If so, we finish the proof. To prove it, we show that

$$\frac{d}{dt}|^{+}\Psi(t) = \limsup_{h\to 0}\frac{1}{h}(\Psi(t+h) - \Psi(t)) \le 0.$$

Notice the fact that

$$|\mathcal{W}_\gamma(\rho_{t+h},\rho) - \mathcal{W}_\gamma(\rho_t,\rho)| \le \mathcal{W}_\gamma(\rho_{t+h},\rho_t).$$

Along the gradient flow $\partial_t\rho = -\mathrm{grad}_g\mathcal{D}_\gamma(\rho\|\mu)$, we have

$$\begin{aligned}\limsup_{h\to 0}\frac{1}{h}\mathcal{W}_\gamma(\rho_{t+h},\rho_t) &= g_\rho(\partial_t\rho_t, \partial_t\rho_t)\\ &= \sqrt{g_\rho(\mathrm{grad}_g\mathcal{D}_\gamma(\rho_t\|\mu), \mathrm{grad}_g\mathcal{D}_\gamma(\rho_t\|\mu))}\\ &= \sqrt{\mathcal{I}_\gamma(\rho_t)}.\end{aligned}$$

In addition, we obtain

$$\begin{aligned}\frac{d}{dt}\sqrt{\frac{2\mathcal{D}_\gamma(\rho_t\|\mu)}{\kappa}} &= \sqrt{\frac{2}{\kappa}\frac{1}{\mathcal{D}_\gamma(\rho_t\|\mu)}}\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu)\\ &= \sqrt{\frac{2}{\kappa}\frac{1}{\mathcal{D}_\gamma(\rho_t\|\mu)}}\Big(-\mathcal{I}_\gamma(\rho_t\|\mu)\Big)\\ &= -\sqrt{\frac{2}{\kappa}\frac{\mathcal{I}_\gamma(\rho_t\|\mu)}{\mathcal{D}_\gamma(\rho_t\|\mu)}}\sqrt{\mathcal{I}_\gamma(\rho_t\|\mu)}\\ &\le -\sqrt{\mathcal{I}_\gamma(\rho_t\|\mu)}.\end{aligned}$$

Thus

$$\begin{aligned}\frac{d}{dt}|^+\Psi(t) &= \limsup_{h\to 0}\frac{\mathcal{W}_\gamma(\rho_{t+h},\rho_0)-\mathcal{W}_\gamma(\rho_t,\rho_0)}{h}+\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu)|_{t=0}\\ &\le \sqrt{\mathcal{I}_\gamma(\rho_t\|\mu)}-\sqrt{\mathcal{I}_\gamma(\rho_t\|\mu)}=0,\end{aligned}$$

which finishes the proof. □

Proof of Theorem 2 We prove an equality by using the Hessian operator of $\mathcal{D}_\gamma(\rho\|\mu)$ in $(\mathcal{P}, g)$ at the point $\rho=\mu$. Notice that for any $\sigma\in T_\rho\mathcal{P}$, then

$$\int \sigma\delta\mathcal{D}_\gamma(\rho\|\mu)dx|_{\rho=\mu}=\int\frac{1}{1-\gamma}(\frac{\rho}{\mu})^{1-\gamma}\sigma dx|_{\rho=\mu}=\frac{1}{1-\gamma}\int\sigma dx=0. \quad (14)$$

We use the Hessian operator formula in (9). Denote $\sigma=-\nabla\cdot(\rho^\gamma\nabla\Phi)$. Then

$$\begin{aligned}\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma)|_{\rho=\mu} &= \int\delta^2\mathcal{D}_\gamma(\rho\|\mu)\sigma^2dx-\int\delta\mathcal{D}_\gamma(\rho\|\mu)\Gamma_\rho(\sigma,\sigma)dx|_{\rho=\mu}\\ &= \int\delta^2\mathcal{D}_\gamma(\rho\|\mu)\sigma^2dx|_{\rho=\mu}\\ &= \int\frac{1}{\mu}\Big(\nabla\cdot(\mu^\gamma\nabla\Phi)\Big)^2dx,\end{aligned}$$

where the second equality uses the fact $\Gamma_\rho(\sigma,\sigma)\in T_\rho\mathcal{P}$ and (14). Comparing the above terms at $\rho=\mu$ in Lemma 11, we prove the equality. □

Proof of Corollary 3 We first prove the following claim.

Claim:

$$
\begin{aligned}
&\min_{\sigma\in T_\rho\mathcal{P}}\Big\{\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma)|_{\rho=\mu}\colon g_\mu(\sigma,\sigma)=1\Big\}\\
&=\min_{\Phi\in C^\infty(M)}\Big\{\int\frac{1}{\mu}\Big(\nabla\cdot(\mu^\gamma\nabla\Phi)\Big)^2dx\colon \int\|\nabla\Phi\|^2\mu^\gamma dx=1\Big\}\\
&=\min_{f\in C^\infty(M)}\Big\{\int\|\nabla f\|^2\mu^\gamma dx\colon \int f^2\mu dx=1,\quad \int f\mu dx=0\Big\}.
\end{aligned}
\tag{15}
$$

Proof of Claim The first equality holds from the definition of Hessian operator at $\rho=\mu$, shown in Theorem 2. We next focus on the second equality. Denote $\sigma_1=-\nabla\cdot(\mu^\gamma\nabla\Phi)$. Then the minimization in the second equation of (15) forms

$$
\lambda_1:=\min_{\sigma_1\in T_\rho\mathcal{P}}\Big\{\int\frac{1}{\mu}\sigma_1^2dx\colon \int(\sigma_1,-\Delta_{\mu^\gamma}^{-1}\sigma_1)dx=1\Big\}.
$$

Its minimizer satisfies the following eigenvalue problem

$$
\frac{1}{\mu}\sigma_1=-\lambda_1\Delta_{\mu^\gamma}^{-1}\sigma_1,
$$

i.e.

$$
-\nabla\cdot(\mu^\gamma\nabla\frac{\sigma_1}{\mu})=\lambda_1\sigma_1.
$$

In other words, $\lambda_1=\lambda_{\min}(-\Delta_{\mu^\gamma}\frac{1}{\mu})$, where $\lambda_{\min}$ represents the smallest non-zero eigenvalue. On the other hand, denote $\sigma_2=f\mu$. Then the minimizer of (15) in the third equality forms

$$
\lambda_2:=\min_{\sigma_2\in T_\rho\mathcal{P}}\Big\{\int\|\nabla\frac{\sigma_2}{\mu}\|^2\mu^\gamma dx\colon \int\frac{\sigma_2^2}{\mu}dx=1\Big\}.
$$

Similarly, the minimizer of above minimization satisfies the following eigenvalue problem

$$
-\frac{1}{\mu}\nabla\cdot(\mu^\gamma\nabla\frac{\sigma_2}{\mu})=\lambda_2\frac{\sigma_2}{\mu}
$$

i.e.

$$
-\nabla\cdot(\mu^\gamma\nabla\frac{\sigma_2}{\mu})=\lambda_2\sigma_2.
$$

Thus $\lambda_2=\lambda_{\min}(-\Delta_{\mu^\gamma}\frac{1}{\mu})$. From the above formulas, we have $\lambda_1=\lambda_2$, which finishes the proof of claim. □

From the above claim, the smallest eigenvalue of Hessian operator of $\mathcal{D}_\gamma$ at $\rho = \mu$ gives the lower bound for Poincare inequality. From the generalized Yano's formula, we have

$$\begin{aligned}
&\mathrm{Hess}_g \mathcal{D}_\gamma(\rho\|\mu)(\sigma,\sigma)|_{\rho=\mu} \\
&= \int \frac{1}{\mu}(-\Delta_{\mu^\gamma}\Phi)^2 dx \\
&= \int \mu^\gamma \Big\{ \big(\mu^{\gamma-1}\mathrm{Ric} - \Delta\mu^{\gamma-1} - \frac{1}{\gamma-1}\mathrm{Hess}\mu^{\gamma-1}\big)(\nabla\Phi,\nabla\Phi) + \mu^{\gamma-1}\|\mathrm{Hess}\Phi\|^2 \\
&\qquad + \gamma(\gamma-1)\mu^{\gamma-1}J(\Phi,\Phi)|_{\rho=\mu}\Big\} dx,
\end{aligned}$$

where

$$\begin{aligned}
J(\Phi,\Phi)|_{\rho=\mu} &= (\nabla\log\mu, \nabla\Phi)^2 - \frac{1}{2}(\frac{\nabla\rho}{\rho}, \frac{\nabla\rho}{\rho} + \frac{\nabla\mu}{\mu})(\nabla\Phi,\nabla\Phi)|_{\rho=\mu} \\
&=(\nabla\log\mu,\nabla\Phi)^2 - \|\nabla\log\mu\|^2\|\nabla\Phi\|^2.
\end{aligned}$$

Thus

$$-\|\nabla\log\mu\|^2\|\nabla\Phi\|^2 \leq J(\Phi,\Phi)|_{\rho=\mu} \leq 0.$$

From the above statement, we can estimate the smallest eigenvalue of Hessian operator, which finishes the proof. □

Proof of Theorem 4 We first prove the $PH^{-1}I$ inequality. Denote ρ_t be a geodesic curve of least energy in $\mathcal{P}$, with H^{-1} metric, where $\rho_0 = \mu$ and $\rho_1 = \rho$. From Proposition 7, $\partial_{tt}\rho_t = 0$, i.e. $\rho_t = (1-t)\rho_0 + t\rho_1$. Thus

$$H^{-1}(\rho,\mu) = \sqrt{(\rho-\mu,\rho-\mu)_{H^{-1}}} = \sqrt{\int(\rho-\mu,(-\Delta)^{-1}(\rho-\mu))dx}.$$

By taking the Taylor expansion of $\mathcal{D}_0(\rho\|\mu)$ in $(\mathcal{P}, H^{-1})$ at $\rho=\mu$, we obtain

$$\begin{aligned}
\mathcal{D}_0(\rho\|\mu) &= \mathcal{D}_0(\mu\|\mu) + (\mathrm{grad}_g\mathcal{D}_0(\rho\|\mu), \rho-\mu)_{H^{-1}} \\
&+ \int (1-t)\mathrm{Hess}_{H^{-1}}\mathcal{F}(\rho_t)(\rho-\mu,\rho-\mu)dt,
\end{aligned} \tag{16}$$

where $\mathcal{D}_0(\mu\|\mu) = 0$. From the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&(\mathrm{grad}_g\mathcal{D}_0(\rho\|\mu), \rho-\mu)_{H^{-1}} \\
&\quad \geq -\sqrt{(\mathrm{grad}_g\mathcal{D}_0(\rho\|\mu), \mathrm{grad}_g\mathcal{D}_0(\rho\|\mu))_{H^{-1}}}\sqrt{(\rho-\mu,\rho-\mu)_{H^{-1}}} \\
&= -\sqrt{\mathcal{I}_0(\rho\|\mu)}H^{-1}(\rho,\mu).
\end{aligned} \tag{17}$$

Hence condition $\mu^{-1}\mathrm{Ric} + \mathrm{Hess}\mu^{-1} - \Delta\mu^{-1} \succeq \kappa$ implies $\mathrm{Hess}_{H^{-1}}\mathcal{D}_0(\rho\|\mu)(\rho - \mu, \rho - \mu) \geq \kappa(\rho - \mu, \rho - \mu)_{H^{-1}}$. Thus

$$\begin{aligned}\int (1-t)\mathrm{Hess}_{H^{-1}}\mathcal{F}(\rho_t)(\rho-\mu,\rho-\mu)dt &\geq \int_0^1 \kappa(1-t)(\rho-\mu,\rho-\mu)_{H^{-1}}dt \\ &= \frac{\kappa}{2}H^{-1}(\rho,\mu)^2.\end{aligned} \tag{18}$$

Substituting (17) and (18) into (16), we prove the $PH^{-1}I$ inequality. In addition, the H^{-1}-Talagrand inequality follows from Theorem 1. □

Remark 9 Our method fails when $\gamma > 1$ or $\gamma < 0$. In these cases, there is no finite lower bound for both bilinear form and squared gradient norm for any $\rho \in \mathcal{P}$. One can not obtain a finite ratio between $\frac{d}{dt}\mathcal{D}_\gamma(\rho_t\|\mu)$ and $\frac{d^2}{dt^2}\mathcal{D}_\gamma(\rho_t\|\mu)$. Thus we can not establish the exponential decay results in term of γ-divergence. However, the current method fails does not mean that we can not find the convergence guarantee condition of γ-diffusion processes when $\gamma > 1$. In fact, we can always formulate γ-divergence as the gradient flow of 1-divergence (relative entropy) w.r.t. density manifold metric $\Big(-\nabla\cdot(\rho\mu^{\gamma-1}\nabla)\Big)^{-1}$. In this case, one can apply a classical Bakry–Emery method. In other words, one can always apply the entropy method or entropy-entropy production as in [29] to find the associated diffusion hypercontractivity and convergence rate in 1-divergence [6].

Remark 10 We remark that the current convergence study is not identical to geodesic convexities in generalized optimal transport spaces [7]. The nonlinear mobilities in Dolbeault–Nazaret–Savare metric spaces bring additional difficulties. The geodesic convexity may not provide the strictly convexity rate for general gradient flows. However, we can still derive convergence rates from Hessian operators along with gradient flows. This requires additional estimations towards the bilinear form J in Lemma 12.

Remark 11 We comment on the proof of different inequalities. (i) For log–Sobolev and Talagrand type inequalities, we only need the Hessian operator along the gradient flow to have a lower bound. (ii) For Poincare inequalities, we require the Hessian operator at the equilibrium measure μ to have a lower bound. (iii) For the divergence, metric and information type inequality, such as HWI or $PH^{-1}I$ inequalities, we require the Hessian operator to have a lower bound for any tangent directions in density manifold. Amazingly, the above three conditions hold when $\gamma = 0, 1$. These choices of parameters correspond to both H^{-1} space and L^2-Wasserstein space.

5 Generalized Bakry–Emery Calculus

In this section, we propose generalized Bakry–Emery iterative calculus. This definition connects Hessian operators in density manifold with generators (Kolmogorov backward operator) of γ-drift–diffusion processes.

We first define generalized iterative Bakry–Emery Gamma operators.

Definition 13 (γ-Bakry–Emery calculus) Denote the γ-Gamma one operator $\Gamma_{\gamma,1}\colon C^\infty(M)\times C^\infty(M)\times\mathcal{P}\to C^\infty(M)$ as

$$\Gamma_{\gamma,1}(\Phi_1,\Phi_2,\rho)=(\nabla\Phi_1,\nabla\Phi_2)\rho^{\gamma-1},$$

where $\Phi_1,\Phi_2\in C^\infty(M)$.

Denote the γ–Gamma two operator $\Gamma_{\gamma,2}\colon C^\infty(M)\times C^\infty(M)\times\mathcal{P}\to C^\infty(M)$ as

$$\begin{aligned}\Gamma_{\gamma,2}(\Phi_1,\Phi_2,\rho)=&\frac{\gamma}{2}L_\gamma\Gamma_{\gamma,1}(\Phi_1,\Phi_2,\rho)-\frac{1}{2}\Gamma_{\gamma,1}(\Phi_1,L_\gamma\Phi_2,\rho)\\&-\frac{1}{2}\Gamma_{\gamma,1}(\Phi_2,L_\gamma\Phi_1,\rho),\end{aligned}$$

where $\Phi_1,\Phi_2\in C^\infty(M)$.

Remark 12 We note that when $\gamma=1$, we reformulate classical iterative Bakry–Emery operators. Note that $\Gamma_{1,1}$ and $\Gamma_{1,2}$ are independent of ρ with

$$\Gamma_{1,1}(\Phi,\Phi)=(\nabla\Phi,\nabla\Phi)\quad\text{and}\quad\Gamma_{1,2}(\Phi,\Phi)=\frac{1}{2}L_1\Gamma_{1,1}(\Phi,\Phi)-\Gamma_{1,1}(\Phi,L_1\Phi),$$

where $L_1=(\nabla\log\mu,\nabla\cdot)+\Delta$ is the generator of a Langevin drift diffusion process. When $\gamma\neq 1$, generalized Bakry–Emery Gamma one and Gamma two operators depend on the current density ρ. In other words, they are mean-field operators.

We next prove an equality to connect generalized Bakry–Emery calculus with Hessian operators of γ-divergences in density manifold.

Proposition 14

$$Hess_g\mathcal{D}_\gamma(\rho\|\mu)(\sigma_1,\sigma_2)=\int\Gamma_{\gamma,2}(\Phi_1,\Phi_2,\rho)(x)\rho(x)dx,$$

where $\sigma_i=-\nabla\cdot(\rho^\gamma\nabla\Phi_i)\in T_\rho\mathcal{P}$, *and* $\Phi_i\in C^\infty(M)$ *with* $i=1,2$.

Proof We omit the notation of ρ with generalized Gamma operators, e.g., $\Gamma_{\gamma,1}(\Phi_1,\Phi_2):=\Gamma_{\gamma,1}(\Phi_1,\Phi_2,\rho)$. We derive the Hessian operator of $\mathcal{D}_\gamma$ in $(\mathcal{P},g)$ by using (9) directly. By using generalized iterative operators, we reformulate (9) as follows:

$$\begin{aligned}&\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(\sigma_1,\sigma_2)\\&\quad=\int\delta^2\mathcal{D}_\gamma\Big(\nabla\cdot(\rho^\gamma\nabla\Phi_1)\Big)\Big(\nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big)dx\\&\quad+\frac{\gamma}{2}\int\Big\{\Gamma_{1,1}(\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma,\Phi_1),\Phi_2)+\Gamma_{1,1}(\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma,\Phi_2),\Phi_1)\\&\qquad\qquad-\Gamma_{1,1}(\Gamma_{\gamma,1}(\Phi_1,\Phi_2),\delta\mathcal{D}_\gamma)\Big\}\rho^\gamma dx.\end{aligned}\tag{19}$$

We next rewrite (19) in three terms. First, we prove the following claim.
Claim 1:

$$\begin{aligned}
&\frac{1}{2}\int \Gamma_{\gamma,1}(\Phi_1, L_\gamma\Phi_2)\rho dx \\
&= \frac{1}{2}\int \delta^2\mathcal{D}_\gamma\Big(\nabla\cdot(\rho^\gamma\nabla\Phi_1)\Big)\Big(\nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big) + \gamma\Gamma_{1,1}(\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma, \Phi_1), \Phi_2)\rho^\gamma dx.
\end{aligned} \tag{20}$$

□

Proof of Claim 1 Notice

$$\int \Gamma_{1,1}(\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma, \Phi_1), \Phi_2)\rho^\gamma dx = -\int \nabla\cdot(\rho^\gamma\nabla\Phi_2)\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma, \Phi_1)dx.$$

and

$$\nabla\cdot(\rho^\gamma\nabla\Phi_1) = (\nabla\rho^\gamma, \nabla\Phi_1) + \rho^\gamma\Delta\Phi_1.$$

The above two facts show that

$$\begin{aligned}
\text{R.H.S. of(20)} &= \frac{1}{2}\int \nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big\{\delta^2\mathcal{D}_\gamma\nabla\cdot(\rho^\gamma\nabla\Phi_1) - \gamma\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma, \Phi_1)\Big\}dx \\
&= \frac{1}{2}\int \nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big\{\delta^2\mathcal{D}_\gamma(\nabla\rho^\gamma, \nabla\Phi_1) \\
&\quad + \delta^2\mathcal{D}_\gamma\rho^\gamma\Delta\Phi_1 - \gamma(\nabla\delta\mathcal{D}_\gamma, \nabla\Phi_1)\rho^{\gamma-1}\Big\}dx.
\end{aligned}$$

Using the fact that $\delta^2\mathcal{D}_\gamma = \rho^{-\gamma}\mu^{\gamma-1}$ and $\nabla\delta\mathcal{D}_\gamma = \rho^{-\gamma}\mu^{\gamma-1}\nabla\rho - \rho^{1-\gamma}\mu^{\gamma-2}\nabla\mu$.

$$\begin{aligned}
\text{R.H.S. of (20)} &= \frac{1}{2}\int \nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big\{\rho^{-\gamma}\mu^{\gamma-1}(\nabla\rho^\gamma, \nabla\Phi_1) + \rho^{-\gamma}\mu^{\gamma-1}\rho^\gamma\Delta\Phi_1 \\
&\quad - \gamma(\rho^{-\gamma}\mu^{\gamma-1}\nabla\rho - \rho^{1-\gamma}\mu^{\gamma-2}\nabla\mu, \nabla\Phi_1)\rho^{\gamma-1}\Big\}dx \\
&= \frac{1}{2}\int \nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big\{\mu^{\gamma-1}\Delta\Phi_1 + \gamma(\mu^{\gamma-2}\nabla\mu, \nabla\Phi_1)\Big\}dx \\
&= \frac{1}{2}\int \nabla\cdot(\rho^\gamma\nabla\Phi_2)L_\gamma\Phi_1 dx \\
&= -\frac{1}{2}\int (\nabla L_\gamma\Phi_1, \nabla\Phi_2)\rho^\gamma dx \\
&= -\frac{1}{2}\int \Gamma_{\gamma,1}(L_\gamma\Phi_1, \Phi_2)\rho dx,
\end{aligned}$$

where the second last equality holds by the integration by parts formula. □

Secondly, by switching Φ_1 and Φ_2 in Claim 1, we have

$$\begin{aligned}
&\frac{1}{2}\int \Gamma_{\gamma,1}(\Phi_2, L_\gamma\Phi_1)\rho dx\\
=&\int \frac{1}{2}\delta^2\mathcal{D}_\gamma\Big(\nabla\cdot(\rho^\gamma\nabla\Phi_1)\Big)\Big(\nabla\cdot(\rho^\gamma\nabla\Phi_2)\Big)+\frac{\gamma}{2}\Gamma_{1,1}(\Gamma_{\gamma,1}(\delta\mathcal{D}_\gamma,\Phi_2),\Phi_1)\rho^\gamma dx.
\end{aligned}\tag{21}$$

Thirdly, we show the following claim.
Claim 2:

$$\int \Gamma_{\gamma,1}(\Phi_1, L_\gamma\Phi_2)\rho dx = \int \Gamma_{1,1}(\Gamma_{\gamma,1}(\Phi_1,\Phi_2),\delta\mathcal{D}_\gamma)\rho^\gamma dx. \tag{22}$$

Proof of Claim 2 Consider

$$\begin{aligned}
\text{R.H.S. of (22)} &= \int \Gamma_{1,1}(\Gamma_{\gamma,1}(\Phi_1,\Phi_2),\delta\mathcal{D}_\gamma)\rho^\gamma dx\\
&= -\int \nabla\cdot(\rho^\gamma\nabla\delta\mathcal{D}_\gamma)\Gamma_{\gamma,1}(\Phi_1,\Phi_2)dx\\
&= -\int L^*_\gamma\rho\Gamma_{\gamma,1}(\Phi_1,\Phi_2)dx\\
&= \int L_\gamma\Gamma_{\gamma,1}(\Phi_1,\Phi_2)\rho dx,
\end{aligned}$$

where the second equality follows

$$\nabla\cdot(\rho^\gamma\nabla\delta\mathcal{D}_\gamma) = \nabla\cdot(\mu^\gamma\nabla\frac{\rho}{\mu}) = L^*_\gamma\rho,$$

and the last equality holds because L^*_γ is the adjoint operator L_γ in $L^2(\rho)$. □

By summing (20), (21) and $\frac{\gamma}{2}$ times (22) and using (19), we have

$$\begin{aligned}
&\mathrm{Hess}_g\mathcal{D}_\gamma(\rho\|\mu)(\sigma_1,\sigma_2)\\
&=\int\Big\{-\frac{1}{2}\Gamma_{\gamma,1}(\Phi_1,L_\gamma\Phi_2)-\frac{1}{2}\Gamma_{\gamma,1}(\Phi_2,L_\gamma\Phi_1)+\frac{\gamma}{2}L_\gamma\Gamma_{\gamma,1}(\Phi_1,\Phi_2)\Big\}\rho(x)dx\\
&=\int \Gamma_{\gamma,2}(\Phi_1,\Phi_2)(x)\rho(x)dx.
\end{aligned}$$

□

We last show that generalized Bakry–Emery iterative calculus implies generalized hypercontractivity.

Proposition 15 (Generalized Bakry–Emery criterion) *If there exists a constant* $\kappa > 0$*, such that*

$$\begin{aligned}
&\int \Gamma_{\gamma,2}(\Phi,\Phi,\rho)(x)\rho(x)dx \geq \kappa\int \Gamma_{\gamma,1}(\Phi,\Phi,\rho)(x)\rho(x)dx,\\
&\Phi = \frac{1}{1-\gamma}(\frac{\rho}{\mu})^{1-\gamma},
\end{aligned}\tag{23}$$

for any $\rho \in \mathcal{P}$. Then the generalized hypercontractivity (2) *and the generalized log–Sobolev inequality* (3) *hold.*

Proof The proof applies Proposition 14 and the gradient flow formulation (10) in deriving Theorem 1. □

Remark 13 Generalized Bakry–Emery operators follow all proofs in [19]. In other words, when the divergence functional is the relative entropy, i.e., $\gamma = 1$, we have the classical Bakry–Emery iterative calculus. To study generalized divergence functionals and drift–diffusion processes, we need to develop generalized iterative Gamma calculus.

Remark 14 When $\gamma = 1$ or $\gamma = 0$, the ratio between generalized Gamma two operator and Gamma one operator provides a bound in (23). This is not the case for $\gamma \neq 1, 0$. In general, we need to apply the mean field (integral formula w.r.t ρ) of the Gamma two operator to bound Gamma one operator. We next derive related log–Sobolev inequalities.

In summary, we show the generalized Bakry–Emery criterion (23), and estimate its bound in Theorem 1. Besides, we comment on major differences between generalized Bakry–Emery criterions and classical ones. The Hessian operator in generalized density manifolds involves an additional quadratic form $J(\Phi, \Phi)$. Thus the smallest eigenvalue of Hessian operator in density manifold is not enough to provide a lower bound for the convergence rate of generalized drift-diffusion processes. We carefully derive the global behavior of dynamics. This is to control the additional quadratic form along with the gradient flow. Besides, a local viewpoint is provided for establishing the Poincare inequality, which is from the Hessian operator in density manifold at the minimizer μ.

In this paper, both Gamma one and Gamma two operators are mean-field operators, which depend on density functions nonlinearly. In future work, we shall study general mean-field Bakry–Emery conditions for related diffusion-diffusion processes and functional inequalities. In addition, we expect to explore interactive studies between optimal transport and information geometry. These studies could be essential in establishing the convergence-guaranteed machine learning sampling algorithms.

Funding W. Li thanks the support of AFOSR MURI FA9550-18-1-0502, AFOSR YIP award No. FA9550-23-1-0087, NSF FRG grant: DMS-2245097, and NSF RTG grant: 2038080.

Data availability Not Applicable.

Declarations

Conflict of interest There is no conflict of interest.

References

1. Amari, S.: Information Geometry and Its Applications, 1st edn. Springer Publishing Company, Incorporated (2016)

2. Amari, S., Cichocki, A.: Information geometry of divergence functions. Bull. Polish Acad. Sci. Tech. Sci. **58**(1), 183–195 (2010)
3. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry, vol. 64. Springer, Cham (2017)
4. Bakry, D., Emery, M.: Diffusions hypercontractives. Seminaire de probabilites de Strasbourg **19**, 177–206 (1985)
5. Bakry, D., Gentil, I., Ledoux, M.: Logarithmic Sobolev Inequalities. In: Analysis and Geometry of Markov Diffusion Operators. Grundlehren der mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics), vol. 348. Springer, Cham (2014)
6. Bolley, F., Gentil, I.: Phi-entropy inequalities for diffusion semigroups. Journal de Mathematiques Pures et Appliquees **93**(5), 449–473 (2010)
7. Carrillo, J.A., Lisini, S., Savare, G., Slepcev, D.: Nonlinear mobility continuity equations and generalized displacement convexity. J. Funct. Anal. **258**(4), 1273–1309 (2010)
8. Cardaliaguet, P., Carlier, G., Nazaret, B.: Geodesics for a class of distances in the space of probability measures. Calc. Var. Partial. Differ. Equ. **48**, 395–420 (2013)
9. Chow, S.N., Li, W., Zhou, H.: Entropy dissipation of Fokker-Planck equations on graphs. Discret. Contin. Dyn. Syst. Ser. A **38**(10), 4929–4950 (2018)
10. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley Series in Telecommunications. Wiley, New York (1991)
11. Csiszar, I., Shields, P.C.: Information theory and statistics: a tutorial. Foundations and Trends™ in Communications and Information Theory **1**(4), 417–528 (2004)
12. Dolbeault, J., Nazaret, B., Savare, G.: A new class of transport distances. Calc. Var. Partial. Differ. Equ. **2**, 193–231 (2010)
13. Dolbeault, J., Nazaret, B., Savare, G.: From Poincare to logarithmic Sobolev inequalities: a gradient flow approach. SIAM J. Math. Anal. (2012)
14. Gross, L.: Logarithmic Sobolev inequalities. Am. J. Math. **97**(4), 1061–1083 (1975)
15. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
16. Juengel, A.: Entropy Methods for Diffusive Partial Differential Equations. Springer Briefs in Mathematics. Springer, New York (2016)
17. Lafferty, J.D.: The density manifold and configuration space quantization. Trans. Am. Math. Soc. **305**(2), 699–741 (1988)
18. Li, W.: A Study on Stochastic Differential Equations and Fokker-Planck Equations with Applications. *phd thesis*, (2016)
19. Li, W.: Transport information geometry: Riemannian calculus on probability simplex. Inf. Geometry **5**, 161–207 (2022)
20. Li, W., Montufar, G.: Ricci curvature for parametric statistics via optimal transport. Inf. Geometry **3**, 89–117 (2020)
21. Li, W., Ying, L.: Hessian transport gradient flows. Res. Math. Sci. **6**, 34 (2019)
22. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. Ann. Math. **169**(3), 903–991 (2009)
23. Monsaingeon, L., Tamanini, L., Vorotnikov, D.: The Dynamical Schrodinger Problem in Abstract Metric Spaces. arXiv:2012.12005, (2020)
24. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Differ. Equ. **26**(1–2), 101–174 (2001)
25. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. **173**(2), 361–400 (2000)
26. Sturm, K.T.: On the geometry of metric measure spaces. Acta Math. **196**(1), 65–131 (2006)
27. Villani, C.: Optimal Transport: Old and New. Number 338 in Grundlehren Der Mathematischen Wissenschaften. Springer, Berlin (2009)
28. Yano, K.: Some remarks on tensor fields and curvature. Ann. Math. 328–347 (1952)
29. Zamponi, N.: *Entropy Methods for Diffusive PDEs. Lecture notes*. (2017)

30. Zozor, S., Brossier, J.M.: deBruijn identities: From Shannon, Kullback-Leibler and Fisher to generalized ϕ-entropies, ϕ-divergences and ϕ-Fisher informations. AIP Conf. Proc. **1641**(1), 522–529 (2015)