HESSIAN METRIC VIA TRANSPORT INFORMATION GEOMETRY

WUCHEN LI

ABSTRACT. We propose to study the Hessian metric of given functional in the space of probability space embedded with L^2–Wasserstein (optimal transport) metric. We name it transport Hessian metric, which contains and extends the classical L^2–Wasserstein metric. We formulate several dynamical systems associated with transport Hessian metrics. Several connections between transport Hessian metrics and math physics equations are discovered. E.g., the transport Hessian gradient flow, including Newton's flow, formulates a mean-field kernel Stein variational gradient flow; The transport Hessian Hamiltonian flow of negative Boltzmann-Shannon entropy forms the Shallow water's equation; The transport Hessian gradient flow of Fisher information forms the heat equation. Several examples and closed-form solutions of finite-dimensional transport Hessian metrics and dynamics are presented.

1. INTRODUCTION

Metrics in probability density space play crucial roles in differential geometry [24, 40], math physics [6, 9, 10, 21, 22, 34, 35], and information theory [13]. One typical application of the metric is Bayesian sampling problem [17], which is to sample a target distribution based on a prior distribution. Here the metric can be used to design objective function. Famous examples include Kullback–Leibler (KL) divergence or Wasserstein distance. Besides the metric can be applied to design sampling directions. Typical examples include Fisher–Rao gradient, transport (Wasserstein) gradient and Stein gradient. These gradient flows in the probability space represent various Markov processes on the sample space, such as birth-death dynamics [1, 3], Langevin dynamics [20], and Stein variational gradient flows [15, 31].

There are naturally geometric variational structures, as well as math physics equations, associated with these metrics. On the one hand, information geometry studies the Hessian metric in probability density space [1, 3]. In this area, the Fisher–Rao metric, i.e. the L^2 Hessian operator of negative Boltzmann–Shannon entropy, plays fundamental roles. It is an invariant metric to derive divergence functions, including the KL divergence. On the other hand, the L^2–Wasserstein metric or transport metric [1] has been studied in the area of optimal transport; see [40, 22] and many references therein. Also, the transport metric deeply interacts with math physics equations. A celebrated result is that the transport gradient flow of negative Boltzmann–Shannon entropy is the heat equation [38, 20]. And

Key words and phrases. Transport Hessian metric; Transport Hessian gradient flows; Transport Hessian Hamiltonian flows; Transport Hessian entropy dissipation; Shallow water equation; Transport Hessian information matrix.

[1]There are many historic names for optimal transport metric [40]. For the simplicity of presentation, we call L^2–Wasserstein metric the *transport metric*.

transport Hamiltonian flows contain compressible Euler equation, Schrödinger equation and Schrödinger bridge problem [10, 12, 21, 24]. Besides, the Stein metrics have been studied independently [31], which formulates transport metrics with given kernel functions. Recently, the Stein metric also finds vast connections with math physics equations and sampling problems [14, 15, 31].

To advance current works, a joint study between information geometry and optimal transport becomes essential. Here the Hessian metrics on density space are of great importance, which arises in differential geometry [8, 39], learning optimization problems [14, 37] and variational formulations of math physics equations [19]. Natural questions arise: *What is the Hessian metric in probability space embedded with the transport metric? What is the formulation of transport Hessian metric related dynamics, such as gradient flows and Hamiltonian flows? Does the transport Hessian dynamics connect with math physics equations?*

In this paper, following studies in [24], we positively answer the above questions. We study the formulation of the Hessian metric for given energy in probability space embedded with optimal transport metric. We name such metric *transport Hessian metric.* We derive both gradient and Hamiltonian flows for the transport Hessian metric. Several observations are made as follows:

(i) The transport Hessian metric is a Stein metric with a mean field kernel. Hence the transport Newton's flow is a particular mean field Stein variational gradient flow;
(ii) At least in one dimensional sample space, the Hamiltonian flow of transport Hessian metric for negative Boltzmann-Shannon entropy forms the Shallow water's equation;
(iii) We establish a relation among negative Boltzmann–Shannon entropy, Fisher information and transport Hessian metric. E.g., the transport Hessian gradient flow of Fisher information satisfies the heat equation.

Our result can be summarized by a variational problem as follows. Given a compact, smooth sample space M, denote the smooth probability density space on M by $\mathcal{P}$ and a smooth energy function $\mathcal{E}\colon \mathcal{P}\to\mathbb{R}$. Consider

$$\begin{aligned}\inf_{\rho,\Phi}\int_0^1\Big\{\int\int &\nabla_x\nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla\Phi(t,x)\nabla\Phi(t,y)\rho(t,x)\rho(t,y)dxdy\\ &+\int\nabla_x^2\delta\mathcal{E}(\rho)(x)(\nabla\Phi(t,x),\nabla\Phi(t,x))\rho(t,x)dx\Big\}dt,\end{aligned}$$

where δ, δ^2 are L^2 first, second variation operators respectively, and the infimum is taken among all possible density paths and potential functions $(\rho,\Phi)\colon[0,1]\times M\to\mathbb{R}^2$, such that the continuity equation with gradient drift vector fields holds

$$\partial_t\rho+\nabla\cdot(\rho\nabla\Phi)=0,\qquad \text{fixed } \rho^0,\ \rho^1.$$

We notice that if $\mathcal{E}(\rho)=\frac{1}{2}\int x^2\rho dx$, then the proposed variational formulation recovers the optimal transport metric [40]. In this sense, the transport Hessian metric contains and extends the L^2–Wasserstein metric.

Here the result belongs to the study of transport information geometry (TIG) founded in [24]. TIG is a newly emerged area, which applies optimal transport, information geometry and differential geometry methods to study variational formulations of math physics equations. Nowadays, it has vast applications in constructing fluid dynamics, proving functional inequalities and designing algorithms; see [7, 16, 23, 25, 26, 27, 28, 29]. In this paper, we extend the area of TIG into the category of Hessian geometry. One direct application is to formulate optimization techniques for Bayesian sampling problems [36, 37]. See related developments in information geometry [32, 33]. In the current paper, we discuss the connections between transport Hessian metrics and math physics equations. We also demonstrate the relation between transport Hessian gradient flows and Stein variational sampling algorithms.

This paper is arranged as follows. Section 2 establishes formulations of transport Hessian metrics and associated distance functions. We derive transport Hessian gradient flows in section 3, whose connections with Stein gradient flows are shown in section 4. In particular, we show that a particular mean field kernel version of Stein variational gradient flow forms the transport Newton's flow. We present Hessian Hamiltonian flows in section 5 and point out the connection between negative Boltzmann–Shanon entropy and Shallow water's equations. Finite dimensional examples of transport Hessian metrics with closed form solutions are presented in section 6.

2. Optimal transport Hessian metric

In this section, we briefly review the definition of the Hessian metric on a finite–dimensional manifold and recall the formulation of the optimal transport metric defined in probability density space. Using these concepts, we introduce the Hessian metric of given energy in probability density space with L^2–Wasserstein metric. Several examples are provided.

2.1. **Finite dimensional Hessian metric.**

Let (M, g) be a smooth, compact, d-dimensional Riemannian manifold without boundaries, where g is the metric tensor of M. For concreteness, let $M = \mathbb{T}^d$, where $\mathbb{T}^d$ represents a d dimensional torus, and assume $g(x) \in C^{d\times d}$ to be a smooth matrix function. In this case, denote ∇ as the Euclidean gradient operator. Here the tangent space and the cotangent space at $x \in M$ form

$$T_xM = \big\{\dot{x} \in \mathbb{R}^d\big\}, \quad T_x^*M = \big\{p = g(x)^{-1}\dot{x} \in \mathbb{R}^d\big\}.$$

Consider an energy function $E\colon M \to \mathbb{R}$. There are several equivalent formulations of Hessian operators defined on (M, g). On the one hand, the Hessian operator of E in (M, g) can be defined on the tangent space, i.e.

$$\mathrm{Hess}_g E\colon M \times T_xM \times T_xM \to \mathbb{R}, \quad \mathrm{Hess}_g E = \Big((\mathrm{Hess}_g E)_{ij}\Big)_{1\le i,j\le d} \in \mathbb{R}^{d\times d}.$$

Here

$$(\mathrm{Hess}_g E(x))_{ij} = (\nabla^2 E(x))_{ij} - \sum_{k=1}^d \nabla_{x_k} E(x)\Gamma_{ij}^k(x),$$

where $\Gamma_{ij}^k\colon M\to\mathbb{R}$ is the Christoffel symbol defined by

$$\Gamma_{ij}^k(x)=\sum_{k'=1}^d (g(x)^{-1})_{kk'}\Big(\nabla_{x_i}g_{jk'}(x)+\nabla_{x_j}g_{ik'}(x)-\nabla_{x'_k}g_{ij}(x)\Big).$$

We note that in our example, Γ_{ij}^k represents the torsion free Levi-Civita connection, i.e. $\Gamma_{ij}^k=\Gamma_{ji}^k$. On the other hand, the Hessian operator can be formulated as a bilinear form on the cotangent space. Denote

$$\mathrm{Hess}_g^*E\colon T_x^*M\times T_x^*M\to\mathbb{R},\quad \mathrm{Hess}_g^*E=\Big((\mathrm{Hess}_g^*E)_{ij}\Big)_{1\leq i,j\leq d}\in\mathbb{R}^{d\times d},$$

then

$$\dot{x}^{\mathsf{T}}\mathrm{Hess}_gE(x)\dot{x}=p^{\mathsf{T}}\mathrm{Hess}_g^*E(x)p,$$

where $p=g(x)^{-1}\dot{x}$, for any $\dot{x}\in T_xM$. In other words, we have

$$\mathrm{Hess}_g^*E(x)=g(x)^{-1}\mathrm{Hess}_gE(x)g(x)^{-1}.$$

Hence the Hessian metric of an energy function E in (M,g) is defined as follows:

$$g^H(x)=\mathrm{Hess}_gE(x)=g(x)\mathrm{Hess}_g^*E(x)g(x).$$

Here (M,g^H) is named Hessian manifold. Later on, we will use the above two formulations of Hessian metric, depending on which is more convenient.

2.2. **Optimal transport metric.** We next present the L^2–Wasserstein metric and demonstrate its associated Hessian operator for a given functional.

Let sample space $(M,g)=(\mathbb{T}^d,\mathbb{I})$ be a d dimensional torus, where $\mathbb{I}\in\mathbb{R}^{d\times d}$ is an identity matrix and $(\cdot,\cdot)$ denotes the Euclidean inner product. Denote a smooth positive probability density space by

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

The tangent space at $\rho\in\mathcal{P}$ is given by

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

To define the optimal transport metric in the probability space, we need the following convention. Define a weighted Laplacian operator by

$$\Delta_a=\nabla\cdot(a\nabla),$$

where $a\in C^\infty(M)$ is a smooth function. Here for any testing functions f_1, $f_2\in C^\infty(M)$, we have

$$\int(f_1,\Delta_af_2)dx=-\int(\nabla f_1,\nabla f_2)adx.$$

We are ready to present the transport metric.

Definition 1 (Transport metric)**.** *The inner product* $\mathbf{g}\colon \mathcal{P}\times T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to\mathbb{R}$ *is defined by*

$$\mathbf{g}(\rho)(\sigma_1,\sigma_2)=\int \big(\sigma_1,(-\Delta_\rho)^{-1}\sigma_2\big)dx,$$

for any $\sigma_1,\sigma_2\in T_\rho\mathcal{P}$*. Here* $\Delta_\rho=-\nabla\cdot(\rho\nabla)\colon C^\infty(M)\times C^\infty(M)$ *is an elliptical operator weighted linearly by density function* ρ*. On the other hand, denote* $\sigma_i=-\Delta_\rho\Phi_i=-\nabla\cdot(\rho\nabla\Phi_i)$*,* $i=1,2$*, then*

$$\begin{aligned}\mathbf{g}(\rho)(\sigma_1,\sigma_2)&=\int \big(\Phi_1,(-\Delta_\rho)(-\Delta_\rho)^{-1}(-\Delta_\rho)\Phi_2\big)dx\\&=\int \big(\Phi_1,-\nabla\cdot(\rho\nabla\Phi_2)\big)dx\\&=\int (\nabla\Phi_1,\nabla\Phi_2)\rho dx.\end{aligned}$$

In this case, $(\mathcal{P},\mathbf{g})$ forms an infinite-dimensional Riemannian manifold, named transport/Wasserstein density manifold [22]. We comment that the optimal transport metric induces a distance function, which has other formulations, such as a linear programming problem with given ground cost function or a mapping formulation, named Monge problems [40]. Throughout this paper, we only use the metric formulation of optimal transport formulated in Definition 1.

We next introduce the Hessian operator in transport density manifold. Denote δ, δ^2 as the L^2 first and second variation operators respectively. Here the tangent space at $\rho\in\mathcal{P}$ forms

$$T_\rho\mathcal{P}=\Big\{\sigma\in C^\infty(M)\Big\},$$

and the cotangent space at $\rho\in\mathcal{P}$ satisfies

$$T_\rho^*\mathcal{P}=\Big\{\Phi=(-\Delta_\rho)^{-1}\sigma\in C^\infty(M)\colon \Phi\text{ is uniquely determined by a constant shrift}\Big\}.$$

Given a smooth functional $\mathcal{E}\colon\mathcal{P}\to\mathbb{R}$, the Hessian operator of $\mathcal{E}$ in $(\mathcal{P},\mathbf{g})$ has the following two formulations. On the one hand, the Hessian operator of $\mathcal{E}$ in $(\mathcal{P},\mathbf{g})$ can be defined on the tangent space, i.e.

$$\mathrm{Hess}_{\mathbf{g}}\mathcal{E}\colon\mathcal{P}\times T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to\mathbb{R},$$

which satisfies

$$\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)(\sigma_1,\sigma_2)=\int\int\delta^2\mathcal{E}(\rho)\sigma_1(x)\sigma_2(y)dxdy+\int\delta\mathcal{E}(\rho)(x)\mathbf{\Gamma}(\rho)(\sigma_1,\sigma_2)(x)dx,\qquad(1)$$

where $\mathbf{\Gamma}\colon\mathcal{P}\times T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to\mathbb{R}$ is the Christoffel symbol in $(\mathcal{P},\mathbf{g})$ defined by

$$\mathbf{\Gamma}(\rho)(\sigma_1,\sigma_2)=-\frac{1}{2}\Big\{\Delta_{\sigma_1}\Delta_\rho^{-1}\sigma_2+\Delta_{\sigma_2}\Delta_\rho^{-1}\sigma_1+\Delta_\rho(\nabla\Delta_\rho^{-1}\sigma_1,\nabla\Delta_\rho^{-1}\sigma_2)\Big\}.\qquad(2)$$

We note here that $\mathbf{\Gamma}$ represents the torsion free Levi-Civita connection in transport density manifold, i.e.

$$\mathbf{\Gamma}(\rho)(\sigma_1,\sigma_2)(x)=\mathbf{\Gamma}(\rho)(\sigma_2,\sigma_1)(x).$$

On the other hand, the Hessian operator can be formulated as a bilinear form on the cotangent space. Denote

$$\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}\colon\mathcal{P}\times T^*_\rho\mathcal{P}\times T^*_\rho\mathcal{P}\to\mathbb{R}.$$

Then by directi calculations, denote

$$\sigma_i = -\nabla\cdot(\rho\nabla\Phi_i), \quad i=1,2,$$

then we have

$$\begin{aligned}\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi_1,\Phi_2) =& \int\int \delta^2\mathcal{E}(\rho)(x,y)(-\Delta_\rho\Phi_1)(x)(-\Delta_\rho\Phi_2)(y)dxdy\\ &-\int \delta\mathcal{E}(\rho)(x)\mathbf{\Gamma}(\rho)(-\Delta_\rho\Phi_1,-\Delta_\rho\Phi_2)(x)dx\\ =&\int\int \nabla_x\nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla\Phi_1(x)\nabla\Phi_2(y)\rho(x)\rho(y)dxdy\\ &+\int \nabla_x^2\delta\mathcal{E}(\rho)(\nabla\Phi_1(x),\nabla\Phi_2(x))\rho(x)dx.\end{aligned} \tag{3}$$

Here the second equality holds by integration by parts formula. In particular, the directly calculations of Christoffel symbol (2) in transport density manifold [24] shows that

$$\begin{aligned}\mathbf{\Gamma}(\rho)(-\Delta_\rho\Phi_1,-\Delta_\rho\Phi_2) =& -\frac{1}{2}\int \delta\mathcal{E}(\rho)(x)\Big\{\nabla\cdot(\nabla\cdot(\rho\nabla\Phi_1)\nabla\Phi_2)+\nabla\cdot(\nabla\cdot(\rho\nabla\Phi_2)\nabla\Phi_1)\\ &+\nabla\cdot(\rho\nabla(\nabla\Phi_1,\nabla\Phi_2))\Big\}dx\\ =&\frac{1}{2}\int \nabla\delta\mathcal{E}(\rho)(x)\Big\{-\nabla\big((\nabla\delta\mathcal{E},\nabla\Phi_1),\nabla\Phi_2\big)-\nabla\big((\nabla\delta\mathcal{E},\nabla\Phi_2),\nabla\Phi_1\big)\\ &+(\nabla\delta\mathcal{E}(\rho),\nabla\big(\nabla\Phi_1,\nabla\Phi_2\big))\Big\}dx\\ =&-\int \nabla_x^2\delta\mathcal{E}(\rho)(\nabla\Phi_1(x),\nabla\Phi_2(x))\rho(x)dx.\end{aligned}$$

Similar to the finite dimensional case, the Hessian operator in density manifold has the following relation:

$$\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho) = \Delta_\rho^{-1}\cdot \mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)\cdot\Delta_\rho^{-1},$$

where operator $\cdot$ is in the sense of L^2 inner product.

2.3. **Optimal transport Hessian metric.** We are now ready to present the transport Hessian metric.

Definition 2 (Transport Hessian metric)**.** *The inner product* $\mathbf{g}^H(\rho)\colon \mathcal{P}\times T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to\mathbb{R}$ *is defined by*

$$\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) = Hess_{\mathbf{g}}\mathcal{E}(\rho)(\sigma_1,\sigma_2),$$

for any $\sigma_1,\sigma_2\in T_\rho\mathcal{P}$. *Here* $Hess_{\mathbf{g}}$ *is defined by* (1). *In details,*

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) =& \int\int \nabla_x\nabla_y\delta^2\mathcal{E}(\rho)(x,y)(\nabla\Phi_1(x),\nabla\Phi_2(y))\rho(x)\rho(y)dxdy\\ &+\int \nabla_x^2\delta\mathcal{E}(\rho)(\nabla\Phi_1(x),\nabla\Phi_2(x))\rho(x)dx,\end{aligned}$$

where

$$\sigma_i = -\nabla\cdot(\rho\nabla\Phi_i), \qquad i=1,2.$$

Here $\mathbf{g}^H_\rho$ is the Hessian metric of energy $\mathcal{E}$ in optimal transport. For this reason, we call $(\mathcal{P}, \mathbf{g}^H(\rho))$ the *Hessian density manifold*. From now on, we only consider the case that $\mathbf{g}^H(\rho)$ is a positive definite operator for all $\rho \in \mathcal{P}$. In other words, we assume that $\mathcal{E}(\rho)$ is strictly geodesically convex in $(\mathcal{P}, \mathbf{g})$.

We next introduce that the transport Hessian metric induces a distance function, $\mathrm{Dist}_H \colon \mathcal{P} \times \mathcal{P} \to \mathbb{R}$. Here Dist_H can be given by the following action functional in $(\mathcal{P}, \mathbf{g}^H(\rho))$:

$$\mathrm{Dist}_H(\rho^0, \rho^1)^2 := \inf_{\rho\colon [0,1]\to\mathcal{P}} \Big\{ \int_0^1 \mathbf{g}^H(\rho)(\partial_t\rho, \partial_t\rho)dt \colon \text{fixed } \rho^0,\ \rho^1 \Big\},$$

where the infimum is taken among all density paths $\rho \colon [0,1] \times M \to \mathbb{R}$. In other words, we arrive at the following definition of distance function.

Definition 3 (Transport Hessian distance)**.**

$$\begin{aligned} Dist_H(\rho^0, \rho^1)^2 = \inf_{\rho,\Phi} \int_0^1 \Big\{ \int\int & \nabla_x\nabla_y \delta^2\mathcal{E}(\rho)(x,y)\nabla\Phi(t,x)\nabla\Phi(t,y)\rho(t,x)\rho(t,y)dxdy \\ & + \int \nabla_x^2 \delta\mathcal{E}(\rho)(\nabla\Phi(t,x), \nabla\Phi(t,x))\rho(t,x)dx \Big\} dt, \end{aligned} \tag{4}$$

where the infimum is taken among all possible density and potential functions $(\rho, \Phi) \colon [0,1]\times M \to \mathbb{R}^2$*, such that the continuity equation with gradient drift vector field holds*

$$\partial_t\rho + \nabla\cdot(\rho\nabla\Phi) = 0, \quad \textit{fixed } \rho^0,\ \rho^1.$$

2.4. **Examples.** We present several examples of transport Hessian metrics. For the simplicity of presentation, several equivalent formats of transport Hessian metrics are given directly; see their derivations in [40] and many references therein. In later on examples, we always denote

$$\Phi_i \in T^*_\rho\mathcal{P} \quad \text{s.t.} \quad \sigma_i = -\nabla\cdot(\rho\nabla\Phi_i) \in T_\rho\mathcal{P}, \quad i = 1, 2.$$

Example 1 (Linear energy)**.** *Consider*

$$\mathcal{E}(\rho) = \int E(x)\rho(x)dx,$$

where $E \in C^\infty(M)$ *is a given strictly convex potential function. Then* $\delta\mathcal{E}(\rho)(x) = E(x)$ *and* $\delta^2\mathcal{E}(\rho)(x,y) = 0$*. Hence*

$$\mathbf{g}^H(\rho)(\sigma_1, \sigma_2) = \int \nabla^2 E(x)(\nabla\Phi_1(x), \nabla\Phi_2(x))\rho(x)dx.$$

In particular, if $E(x) = \frac{x^2}{2}$*, then*

$$\mathbf{g}^H(\rho)(\sigma_1, \sigma_2) = \int(\nabla\Phi_1(x), \nabla\Phi_2(x))\rho(x)dx.$$

In this case, our transport Hessian metric is exactly the L^2*–Wasserstein metric; see* [40]*. Hence the transport Hessian metric contains and extends the formulation of optimal transport metric.*

Remark 1. We remark that the L^2–Wasserstein metric itself is a Hessian metric, which is a Hessian operator of $\mathcal{E}(\rho) = \frac{1}{2}\int x^2\rho dx$ in transport density manifold $(\mathcal{P}, \mathbf{g})$.

Example 2 (Interaction energy)**.** *Consider*

$$\mathcal{E}(\rho) = \frac{1}{2}\int\int W(x,y)\rho(x)\rho(y)dxdy,$$

where $W \in C^\infty(M\times M)$ *is a given kernel potential function. Then* $\delta\mathcal{E}(\rho)(x) = \int W(x,y)\rho(y)dy$ *and* $\delta^2\mathcal{E}(\rho)(x,y) = W(x,y)$. *Hence*

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) = \int\int\Big[&\nabla^2_{xy}W(x,y)(\nabla_x\Phi_1(x),\nabla_y\Phi_2(y))\\ &+\nabla^2_x W(x,y)(\nabla\Phi_1(x),\nabla\Phi_2(x))\Big]\rho(x)\rho(y)dxdy.\end{aligned}$$

A concrete examples is given as follows: If $W(x,y) = \frac{\|x-y\|^2}{2}$, *then* $-\nabla^2_{xy}W(x,y) = \nabla^2_x W(x,y) = \mathbb{I}$. *Hence*

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) &= \int(\nabla\Phi_1(x),\nabla\Phi_2(x))\rho dx - \Big(\int\nabla\Phi_1(y)\rho(y)dy, \int\nabla\Phi_2(y)\rho(y)dy\Big)\\ &= \int\Big(\nabla\Phi_1(x) - \int\nabla\Phi_1(y)\rho(y)dy, \nabla\Phi_2(x) - \int\nabla\Phi_2(y)\rho(y)dy\Big)\rho(x)dx.\end{aligned}$$

Example 3 (Entropy)**.** *Consider*

$$\mathcal{E}(\rho) = \int f(\rho)(x)dx,$$

where $f \in C^\infty(\mathbb{R})$ *is a given strict convex function. In this case,* $\delta_\rho\mathcal{E}(\rho)(x) = f'(\rho)(x)$ *and* $\delta^2\mathcal{E}(\rho)(x,y) = f''(\rho)\delta(x-y)$, *where* $\delta(x-y)$ *is a delta function. Hence*

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma,\sigma) &= \int\Big\{f''(\rho)(x)\nabla\cdot(\rho(x)\nabla\Phi_1(x))\nabla\cdot(\rho(x)\nabla\Phi_2(x))\\ &\qquad+\nabla^2 f'(\rho)(x)(\nabla\Phi_1(x),\nabla\Phi_2(x))\rho(x)\Big\}dx \qquad (5)\\ &= \int tr(\nabla^2\Phi_1(x),\nabla^2\Phi_2(x))p(\rho)(x) + (\Delta\Phi_1(x),\Delta\Phi_2(x))p_2(\rho)(x)dx,\end{aligned}$$

where tr denotes the matrix trace operator, and functions p, $p_2\colon\mathbb{R}\to\mathbb{R}$ *satisfy*

$$p(\rho) = \rho f'(\rho) - f(\rho), \quad p_2(\rho) = \rho p'(\rho) - p(\rho).$$

Several concrete examples are provided as follows:

(i) *If* $f(\rho) = \rho\log\rho$, *then* $\mathcal{E}(\rho) = -\int\rho\log\rho dx$ *is known as the negative Boltzmann–Shannon entropy. And* $f'(\rho) = \log\rho + 1$, $f''(\rho) = \frac{1}{\rho}$, *hence* $p(\rho) = \rho$, $p_2(\rho) = 0$. *Then*

$$\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) = \int tr(\nabla^2\Phi_1(x),\nabla^2\Phi_2(x))\rho(x)dx. \qquad (6)$$

(ii) *If* $f(\rho) = \frac{1}{2}\rho^2$, *then* $\mathcal{E}(\rho) = \frac{1}{2}\int\rho^2 dx$. *And* $f'(\rho) = \rho$, $f''(\rho) = 1$. *Hence* $p(\rho) = p_2(\rho) = \frac{1}{2}\rho^2$. *Then*

$$\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) = \frac{1}{2}\int\Big(tr(\nabla^2\Phi_1(x),\nabla^2\Phi_2(x)) + (\Delta\Phi_1(x),\Delta\Phi_2(x))\Big)\rho(x)^2dx.$$

Remark 2. If (M, g) is a Riemannian manifold, then Hessian operator in (5) also contains the Ricci curvature tensor on (M, g). For the simplicity of presentation, we assume that manifold (M, g) is Ricci flat, i.e. $\mathrm{Ric}_M = 0$.

Remark 3. We notice that the second equation in formula (5) has been formulated in [40]; where the first equality equals to the second equality can be shown by using Wasserstein Christoffel symbol (2); see [24, 25]. This fact relates to the construction of Bakry–Émery Gamma calculus; see related works in [5, 16, 25].

Remark 4. We remark that in one dimensional sample space, the transport Hessian metric defined in (6) coincides with the semi–invariant metric studied in [4].

3. Transport Hessian gradient flows

In this section, we derive gradient flows in the Hessian density manifold and provide the associated entropy dissipation property. Several examples are given.

We first derive the gradient flow in Hessian density manifold $(\mathcal{P}, \mathbf{g}^H)$. From now on, we consider a smooth energy $\mathcal{F} \colon \mathcal{P} \to \mathbb{R}$.

Theorem 4 (Transport Hessian gradient flow)**.** *The gradient flow of $\mathcal{F}(\rho)$ in $(\mathcal{P}, \mathbf{g}^H)$ satisfies*

$$\partial_t \rho(t,x) = \nabla \cdot (\rho(t,x) \nabla \Phi^{\mathcal{F}}(x, \rho)), \tag{7}$$

where $\Phi^{\mathcal{F}} \in C^\infty(M \times \mathcal{P})$ satisfies the following Poisson equation:

$$\begin{aligned}
&\nabla_x \cdot (\rho(t,x) \nabla_x \int \nabla_y \delta^2 \mathcal{E}(\rho)(x,y) \nabla_y \Phi^{\mathcal{F}}(y,\rho) \rho(t,y) dy) + \nabla_x \cdot (\rho(t,x) \nabla_x^2 \delta \mathcal{E}(\rho)(x) \nabla_x \Phi^{\mathcal{F}}(x,\rho)) \\
=&\nabla_x \cdot (\rho(t,x) \nabla_x \delta \mathcal{F}(\rho)(x)).
\end{aligned} \tag{8}$$

Proof. The proof follows the definition of gradient operator.

Claim: The gradient operator of $\mathcal{F}(\rho)$ in $(\mathcal{P}, \mathbf{g}^H)$, i.e. $\mathrm{grad}_{\mathbf{g}^H} \colon \mathcal{P} \times C^\infty(\mathcal{P}) \to T_\rho \mathcal{P}$, is defined by

$$\mathrm{grad}_{\mathbf{g}^H} \mathcal{F}(\rho)(x) = -\nabla \cdot (\rho(x) \nabla \Phi^{\mathcal{F}}(x, \rho)),$$

where $\Phi^{\mathcal{F}} \in C^\infty(M)$ solves Poisson equation (8).

Suppose the claim is true, then the gradient flow follows

$$\partial_t \rho = -\mathrm{grad}_{\mathbf{g}^H} \mathcal{F}(\rho),$$

which finishes the proof. Now, we only need to prove the claim as follows.

Proof of claim. Here for any $\sigma(x) \in T_\rho \mathcal{P}$, we have

$$\mathbf{g}^H(\rho)(\sigma, \mathrm{grad}_{\mathbf{g}^H} \mathcal{F}(\rho)) = \int \delta \mathcal{F}(\rho)(x) \sigma(x) dx.$$

Here denote $\Phi,\ \Phi^{\mathcal{F}} \in C^\infty(M \times \mathcal{P})$, such that

$$\sigma = -\nabla \cdot (\rho \nabla \Phi), \quad \text{and} \quad \mathrm{grad}_{\mathbf{g}^H} \mathcal{F}(\rho) = -\nabla \cdot (\rho \nabla \Phi^{\mathcal{F}}).$$

Then from (3), we obtain

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma, \mathrm{grad}_{\mathbf{g}^H}\mathcal{F}(\rho)) =&\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)(-\nabla\cdot(\rho\nabla\Phi), -\nabla\cdot(\rho\nabla\Phi^{\mathcal{F}}))\\ =&\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi, \Phi^{\mathcal{F}})\\ =&\int\int \nabla_x\nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla_x\Phi(x)\nabla_y\Phi^{\mathcal{F}}(y,\rho)\rho(x)\rho(y)dxdy\\ &+\int\nabla_x^2\delta\mathcal{E}(\rho)(\nabla\Phi(x), \nabla\Phi^{\mathcal{F}}(x,\rho))\rho(x)dx\\ =&-\int\nabla_x\cdot\Big(\rho(x)\nabla_x\int\nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla\Phi^{\mathcal{F}}(y,\rho)\rho(y)dy\Big)\Phi(x)dx\\ &-\int\nabla_x\cdot(\rho\nabla_x^2\delta\mathcal{E}(\rho)\nabla\Phi^{\mathcal{F}}(x,\rho))\Phi(x)dx,\end{aligned}$$

where the last equality follows from the integration by parts formula. And

$$\begin{aligned}\int\delta\mathcal{F}(\rho)(x)\sigma(x)dx =& \int\delta\mathcal{F}(\rho)(x)\Big(-\nabla\cdot(\rho\nabla\Phi)(x)\Big)dx\\ =&-\int\Big(\nabla\cdot(\rho\nabla\delta\mathcal{F}(\rho)(x))\Big)\Phi(x)dx,\end{aligned}$$

where the last equality holds by the integration by parts formula twice. We notice that the above two formulas equal to each other for any $\sigma\in T_\rho\mathcal{P}$, i.e. for any $\Phi\in C^\infty(M)$. This derives the Poisson equation (8). □

□

We next present the following two categories of gradient flows in Hessian density manifold. Firstly, we introduce a class of transport Newton's flows [36].

Corollary 5 (Transport Newton's flow)**.** *If $\mathcal{E}(\rho)=\mathcal{F}(\rho)$, then the gradient flow of $\mathcal{F}(\rho)$ in Hessian density manifold $(\mathcal{P}, \mathbf{g}^H)$ forms*

$$\begin{cases}\partial_t\rho(t,x)=\nabla\cdot(\rho(t,x)\nabla\Phi(x,\rho))\\ \nabla_x\cdot(\rho(t,x)\nabla_x\int\nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla_y\Phi(y,\rho)\rho(t,y)dy)+\nabla_x\cdot(\rho(t,x)\nabla_x^2\delta\mathcal{E}(\rho)(x)\nabla_x\Phi(x,\rho))\\ =\nabla\cdot(\rho(t,x)\nabla\delta\mathcal{E}(\rho)(x)).\end{cases}\tag{9}$$

This is the Newton's flow of $\mathcal{E}(\rho)$ in density manifold $(\mathcal{P}, \mathbf{g})$.

Remark 5. We comment that the Newton's flow in transport density manifold and the Newton's flow in L^2 space behave similarly in the asymptotical sense. In other words, we observe that when the density is at the minimizer, i.e. $\rho=\rho^*$, then $\nabla_x\delta\mathcal{E}(\rho^*)=0$. Here the transport Newton's direction forms the L^2 Newton's direction, i.e.

$$\mathrm{grad}_{\mathbf{g}^H}\mathcal{E}(\rho)|_{\rho=\rho^*}=\delta^2\mathcal{E}(\rho)^{-1}\delta\mathcal{E}(\rho)|_{\rho=\rho^*},$$

This fact demonstrates that the transport Newton's direction is asymptotically a L^2–Newton's direction in density space. See related convergence proof of transport Newton's method in [37].

Proof. The proof follows from the definition of Newton's direction in a manifold. Notice that the Newton's flow forms

$$\begin{aligned}\partial_t\rho =& -\Big(\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)\Big)^{-1}\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho)\\ =& -\mathrm{grad}_{\mathbf{g}^H}\mathcal{E}(\rho).\end{aligned}$$

Substituting $\mathcal{F}(\rho)=\mathcal{E}(\rho)$ into equation (7), we obtain the result. □

Secondly, we propose a new class of metrics for deriving transport/Wasserstein gradient flows.

Corollary 6 (Transport gradient flow)**.** *Consider an energy*

$$\mathcal{F}(\rho)=\frac{1}{2}\int\|\nabla_x\delta\mathcal{E}(\rho)(x)\|^2\rho(x)dx.$$

Then the gradient flow of energy $\mathcal{F}(\rho)$ in Hessian density manifold $(\mathcal{P},\mathbf{g}^H)$ formulates

$$\partial_t\rho(t,x)=\nabla_x\cdot(\rho(t,x)\nabla_x\delta\mathcal{E}(\rho)(x)),$$

which is the gradient flow of energy $\mathcal{E}(\rho)$ in density manifold $(\mathcal{P},\mathbf{g})$.

Proof. We notice that

$$\begin{aligned}\mathcal{F}(\rho)=&\frac{1}{2}\int\Big(\delta\mathcal{E}(\rho),-\Delta_\rho\delta\mathcal{E}(\rho)\Big)dx\\ =&\frac{1}{2}\int\|\nabla_x\delta\mathcal{E}(\rho)(x)\|^2\rho(x)dx\\ =&\frac{1}{2}\mathbf{g}(\rho)(\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho),\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho)).\end{aligned}$$

Then

$$\begin{aligned}\mathrm{grad}_{\mathbf{g}}\mathcal{F}(\rho)=&\mathrm{grad}_{\mathbf{g}}\Big\{\frac{1}{2}\mathbf{g}(\rho)(\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho),\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho))\Big\}\\ =&\mathrm{Hess}_g\mathcal{E}(\rho)\cdot\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho).\end{aligned}$$

Hence the gradient flow equation of $\mathcal{F}(\rho)$ in $(\mathcal{P},\mathbf{g}^H)$ satisfies

$$\begin{aligned}\partial_t\rho=&-\mathrm{grad}_{\mathbf{g}^H}\mathcal{F}(\rho)\\ =&-\Big(\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)\Big)^{-1}\mathrm{grad}_{\mathbf{g}}\mathcal{F}(\rho)\\ =&-\Big(\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)\Big)^{-1}\cdot\mathrm{Hess}_g\mathcal{E}(\rho)\cdot\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho)\\ =&-\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho)\\ =&\nabla\cdot(\rho\nabla\delta\mathcal{E}(\rho)),\end{aligned}$$

which finishes the proof. □

3.1. **Examples.** In this section, we present several examples of gradient flows in Hessian density manifold, and points out their connections with math physics equations.

We first present several examples of transport Newton's flows, in which we consider the energy functional $\mathcal{F}(\rho) = \mathcal{E}(\rho)$.

Example 4 (Transport Newton's flow of linear energy)**.** *Consider*

$$\mathcal{E}(\rho) = \int E(x)\rho(x)dx,$$

then the transport Newton's flow forms

$$\begin{cases} \partial_t\rho(t,x) = \nabla\cdot(\rho(t,x)\nabla\Phi^{\mathcal{E}}(x,\rho)) \\ \nabla\cdot(\rho(t,x)\nabla^2E(x)\nabla\Phi^{\mathcal{E}}(x,\rho)) = \nabla\cdot(\rho(t,x)\nabla E(x)). \end{cases}$$

Example 5 (Transport Newton's flow of interaction energy)**.** *Consider*

$$\mathcal{E}(\rho) = \frac{1}{2}\int\int W(x,y)\rho(x)\rho(y)dxdy,$$

then the transport Newton's flow satisfies

$$\begin{cases} \partial_t\rho(t,x) = \nabla\cdot(\rho(t,x)\nabla\Phi^{\mathcal{E}}(x,\rho)) \\ \nabla\cdot\Big(\rho(t,x)\big[\int\nabla^2_{xy}W(x,y)\nabla\Phi^{\mathcal{E}}(y,\rho)\rho(t,y)dy + \int\nabla^2_xW(x,y)\nabla\Phi^{\mathcal{E}}(x,\rho)\rho(t,y)dy\big]\Big) \\ = \nabla\cdot(\rho(t,x)\nabla\int W(x,y)\rho(t,y)dy). \end{cases}$$

A concrete example of interaction kernel is given as follows. If $W(x,y) = \frac{\|x-y\|^2}{2}$*, then we obtain*

$$\begin{cases} \partial_t\rho(t,x) = \nabla\cdot(\rho(t,x)\nabla\Phi^{\mathcal{E}}(x,\rho)) \\ \nabla\cdot\Big(\rho(t,x)\big[\nabla_x\Phi^{\mathcal{E}}(x,\rho) - \int\nabla_y\Phi^{\mathcal{E}}(y,\rho)\rho(t,y)dy\big]\Big) = \nabla\cdot(\rho(t,x)\nabla\int\frac{1}{2}\|x-y\|^2\rho(t,y)dy). \end{cases}$$

Example 6 (Transport Newton's flow of entropy)**.** *Consider*

$$\mathcal{E}(\rho) = \int f(\rho)(x)dx,$$

then the transport Newton's flow satisfies

$$\begin{cases} \partial_t\rho(t,x) = \nabla\cdot(\rho(t,x)\nabla\Phi^{\mathcal{E}}(x,\rho)) \\ -\Big\{\nabla^2\colon(p(\rho(t,x))\nabla^2\Phi^{\mathcal{E}}(x,\rho)) + \Delta(p_2(\rho(t,x))\Delta\Phi^{\mathcal{E}}(x,\rho))\Big\} = \nabla\cdot(\rho(t,x)\nabla f'(\rho)(x)). \end{cases} \tag{10}$$

Here for any function $a\in C^\infty(M)$*,* $\nabla^2\colon(a\nabla^2)$ *represents the second order weighted Laplacian operator. In other words, for any testing function* $f_1,\ f_2\in C^\infty(M)$*, we have*

$$\int f_1\nabla^2\colon(a\nabla^2f_2)dx = \int tr(\nabla^2f_1,\nabla^2f_2)adx.$$

Several concrete examples of equation (10) *are given as follows.*

(i) *If* $f(\rho)=\rho\log\rho$, *then we have*

$$\begin{cases}\partial_t\rho(t,x)=\nabla\cdot(\rho(t,x)\nabla\Phi^{\mathcal{E}}(x,\rho))\\ -\nabla^2\colon(\rho(t,x)\nabla^2\Phi^{\mathcal{E}}(x,\rho))=\nabla\cdot(\rho(t,x)\nabla\log\rho(t,x))=\Delta\rho(t,x).\end{cases}$$

(ii) *If* $f(\rho)=\frac{\rho^2}{2}$, *then we obtain*

$$\begin{cases}\partial_t\rho(t,x)=\nabla\cdot(\rho(t,x)\nabla\Phi^{\mathcal{E}}(x,\rho))\\ -\frac{1}{2}\nabla^2\colon(\rho(t,x)^2\nabla^2\Phi^{\mathcal{E}}(x,\rho))-\frac{1}{2}\Delta(\rho(t,x)^2\Delta\Phi^{\mathcal{E}}(x,\rho))\\ =\nabla\cdot(\rho(t,x)\nabla\rho(t,x))=\frac{1}{2}\Delta\rho^2(t,x).\end{cases}$$

We notice that equation (10) *introduces transport Newton's equations for general entropy functions. Notice that the transport gradient flows of entropy functions include the heat equation, Porous media equation* [40]*, etc. Following this relation, we named the derived equations as* Newton's heat equation, Newton's Porous media equation*, etc.*

We next present the other connection between transport Hessian metrics and heat equations.

Example 7 (Connections with heat equations)**.** *Consider* $\mathcal{E}(\rho)$ *as the negative Boltzmann–Shannon entropy*

$$\mathcal{E}(\rho)=\int\rho(x)\log\rho(x)dx.$$

Denote $\mathcal{F}(\rho)$ *as the* $1/2$ *Fisher information functional*

$$\mathcal{F}(\rho)=\frac{1}{2}\mathbf{g}(\rho)(grad_{\mathbf{g}}\mathcal{E}(\rho),grad_{\mathbf{g}}\mathcal{E}(\rho))=\frac{1}{2}\int\|\nabla\log\rho\|^2\rho dx.$$

The the transport Hessian gradient flow of $1/2$ *Fisher information function forms the heat equation.*

$$\begin{aligned}\partial_t\rho=&-grad_{\mathbf{g}^H}\mathcal{F}(\rho)\\ =&-\nabla\cdot(\rho\nabla)(\nabla^2\colon\rho\nabla^2)^{-1}\nabla\cdot(\rho\nabla)\delta_\rho\mathcal{F}(\rho)\\ =&\nabla\cdot(\rho\nabla)(\nabla^2\colon\rho\nabla^2)^{-1}(-\nabla^2\colon\rho\nabla^2)\delta\mathcal{E}(\rho)\\ =&\nabla\cdot(\rho\nabla\delta\mathcal{E}(\rho))=\nabla\cdot(\rho\nabla\log\rho)\\ =&\Delta\rho.\end{aligned}$$

In above, we use the fact [18] *that the gradient operator of Fisher information functional in* $(\mathcal{P},\mathbf{g})$ *satisfies*

$$grad_{\mathbf{g}}\mathcal{F}(\rho)=Hess_{\mathbf{g}}\mathcal{E}(\rho)\cdot grad_{\mathbf{g}}\mathcal{E}(\rho)=-\nabla^2\colon(\rho\nabla^2\log\rho).$$

Remark 6. We notice that the gradient flow of Fisher information in L^2–Wasserstein metric is known as the quantum heat equation; see [18] and many references therein. We summarize that the relation among heat equation, quantum heat equation and Newton's heat equation is as follows. Denote $\mathcal{E}(\rho)=\int \rho\log\rho dx$ and $\mathcal{F}(\rho)=\frac{1}{2}\int\|\nabla\log\rho\|^2\rho dx$, then

$$\partial_t\rho = -\mathrm{grad}_{\mathbf{g}}\mathcal{F}(\rho) = -\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)\cdot\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho); \qquad \text{Quantum heat equation}$$

$$\partial_t\rho = -\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)^{-1}\cdot\mathrm{grad}_{\mathbf{g}}\mathcal{F}(\rho) = -\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho); \qquad \text{Heat equation.}$$

$$\partial_t\rho = -\mathrm{Hess}_{\mathbf{g}}\mathcal{E}(\rho)^{-1}\cdot\mathrm{grad}_{\mathbf{g}}\mathcal{E}(\rho); \qquad \text{Newton's heat equation}$$

3.2. **Transport Hessian entropy dissipation.** In this subsection, we demonstrate the dissipation property of energy functional along transport Hessian gradient flows.

Corollary 7 (Transport Hessian de Brun identity)**.** *Suppose $\rho(t,x)$ satisfies the transport Hessian gradient flow* (7)*, then*

$$\frac{d}{dt}\mathcal{F}(\rho(t,\cdot)) = -\,\mathcal{I}_H(\rho(t,\cdot)),$$

where $\mathcal{I}_H\colon\mathcal{P}\to\mathbb{R}$ is defined by

$$\mathcal{I}_H(\rho)=\int(\nabla_x\Phi^{\mathcal{F}}(x,\rho),\nabla_x\delta\mathcal{F}(\rho)(x))\rho(x)dx,$$

with $\Phi^{\mathcal{F}}$ satisfying the Poisson equation (8)*.*

Proof. The proof follows from the definition of gradient flow. In other words,

$$\begin{aligned}\frac{d}{dt}\mathcal{F}(\rho) &= \int\delta\mathcal{F}(\rho)(x)\nabla_x\cdot(\rho(t,x)\nabla_x\Phi^{\mathcal{F}}(x,\rho))dx\\ &= -\int(\nabla_x\delta\mathcal{F}(\rho)(x),\nabla_x\Phi^{\mathcal{F}}(x,\rho))\rho(t,x)dx,\end{aligned}$$

which finishes the proof. □

We provide several examples of the proposed entropy dissipation relations.

Example 8. *We notice that if $\mathcal{E}(\rho)=\frac{1}{2}\int x^2\rho(x)dx$ and $\mathcal{F}(\rho)=\int\rho(x)\log\rho(x)dx$, then*

$$\nabla\cdot(\rho\nabla\Phi^{\mathcal{F}}(x,\rho))=\nabla\cdot(\rho\nabla\log\rho),$$

i.e.

$$\Phi^{\mathcal{F}}(x,\rho)=\log\rho.$$

Hence

$$\mathcal{I}_H(\rho)=\int\|\nabla\log\rho(x)\|^2\rho(x)dx,$$

which is known as the Fisher information functional. Hence, for general choices of $\mathcal{E}$, functional $\mathcal{I}_H$ extends the definition of classical Fisher information functional. For this reason, we call $\mathcal{I}_H$ the transport Hessian Fisher information functional*.*

Example 9. *If $\mathcal{E}(\rho)=\int\rho(x)\log\rho(x)dx$ and $\mathcal{F}(\rho)=\frac{1}{2}\int\|\nabla\log\rho(x)\|^2\rho(x)dx$, then*

$$\nabla^2\colon(\rho\nabla^2\Phi^{\mathcal{F}}(x,\rho))=-\nabla\cdot(\rho\nabla\delta\mathcal{F}(\rho))=\nabla^2\colon(\rho\nabla^2\log\rho),$$

i.e.

$$\Phi^{\mathcal{F}}(x,\rho)=\log\rho.$$

Hence

$$\mathcal{I}_H(\rho)=\int tr(\nabla^2\log\rho(x),\nabla^2\log\rho(x))\rho(x)dx.$$

In this case, our transport Hessian Fisher information functional recovers the second order entropy dissipation property. In other words, the dissipation of classical Fisher information functional $\int \|\nabla \log\rho\|^2\rho dx$ along heat equation equals to second order information functional $\mathcal{I}_H = \int tr(\nabla^2\log\rho, \nabla^2\log\rho)\rho dx$; see [39].

We notice that this entropy dissipation relation will be useful in proving related functional inequalities. We leave the related studies of functional inequalities in the future.

4. Connections with Stein variational gradient flows

In this section, we focus on the connection between transport Hessian gradient flows and Stein variational gradient flows.

To do so, we first introduce the following definition of kernel functions.

Definition 8. *Denote the kernel functions H, $K_H\colon M\times M\times\mathcal{P}\to\mathbb{R}$ as follows:*

$$H(x,y,\rho) = \delta^2\mathcal{E}(\rho)(x,y)\nabla_x\cdot(\rho\nabla_x)\nabla_y\cdot(\rho\nabla_y) - \nabla_x\cdot(\rho\nabla_x^2\delta\mathcal{E}(\rho)(x)\nabla_x)\delta(x-y). \quad (11)$$

Denote $K_H(x,y,\rho) := H(x,y,\rho)^{-1}$, in the sense that

$$\int\int K_H(x,y,\rho)H(y,z,\rho)f(z)dxdy = f(z),$$

for any testing function $f\in C^\infty(M)$.

Using the kernel function K_H, we formulate the transport Hessian metric as follows.

Theorem 9. *Given $\sigma_i\in T_\rho\mathcal{P}$, $i=1,2$, the transport Hessian metric $\mathbf{g}^H$ satisfies*

$$\mathbf{g}^H(\rho)(\sigma_1,\sigma_2) = \int\int \nabla_x\nabla_y K_H(x,y,\rho)\big(\nabla_x\Psi_1(x),\nabla_y\Psi_2(y)\big)\rho(x)\rho(y)dxdy, \quad (12)$$

where $\Psi_i\in C^\infty(M)$ satisfies

$$\sigma_i(x) = -\nabla_x\cdot\Big(\rho(x)\int\nabla_x\nabla_y K_H(x,y,\rho)\nabla_y\Psi_i(y)\rho(y)dy\Big). \quad (13)$$

Remark 7. We remark that the transport Hessian metric is a particular Stein variational metric [15, 31] with a mean field kernel function. To see this connection, let us first recall the definition of a Stein metric. Given a kernel matrix function $K_S(x,y) = K_S(y,x)\in\mathbb{R}^{d\times d}$, for any $x,y\in M$ with suitable conditions, the Stein metric $\mathbf{g}^S\colon\mathcal{P}\times T_\rho\mathcal{P}\times T_\rho\mathcal{P}\to\mathbb{R}$ is defined as follows:

$$\mathbf{g}^S(\rho)(\sigma_1,\sigma_2) = \int\int K_S(x,y)(\nabla_x\Psi(x),\nabla_y\Psi(y))\rho(x)\rho(y)dxdy, \quad (14)$$

with

$$\sigma_i(x) = -\nabla_x\cdot(\rho(x)\int\rho(y)K_S(x,y)\nabla_y\Psi_i(y)dy), \quad \text{for } i=1,2.$$

Comparing the definition of Stein metric (14) with transport Hessian metric (12), we notice that

$$K_S(x,y) = \nabla_x\nabla_y K_H(x,y,\rho).$$

Here the matrix kernel function K_S in transport Hessian metric depends on current density function ρ.

Proof. We notice that

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma,\sigma) =&\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi,\Phi)\\ =&\int\int H(x,y,\rho)\Phi(x)\Phi(y)dxdy,\end{aligned}$$

where

$$\sigma=-\nabla\cdot(\rho\nabla\Phi).$$

And $H\colon M\times M\times\mathcal{P}\to\mathbb{R}$ is the kernel function satisfying (11). Denote $\Phi=(-\Delta_\rho)^{-1}\sigma$, then

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma,\sigma)=&\int\int H(x,y,\rho)(-\Delta_\rho)^{-1}\sigma(x)(-\Delta_\rho)^{-1}\sigma(y)dxdy\\ =&\Big(\sigma,(-\Delta_\rho)^{-1}\cdot H\cdot(-\Delta_\rho)^{-1}\sigma\Big)_{L^2}.\end{aligned}$$

We next represent the metric $\mathbf{g^H}$ in the cotangent space of $(\mathcal{P},\mathbf{g}^H)$. Denote

$$\begin{aligned}\sigma=&\Big((-\Delta_\rho)^{-1}\cdot H\cdot(-\Delta_\rho)^{-1}\Big)^{-1}\Psi\\ =&\Delta_\rho\cdot H^{-1}\cdot\Delta_\rho\Psi\\ =&\Delta_\rho\cdot K_H\cdot\Delta_\rho\Psi.\end{aligned}$$

By integration by parts formula, we derive equation (13). Denote the operator $G_H\colon T_\rho\mathcal{P}\to T_\rho\mathcal{P}$ by

$$G_H(\rho)=(\Delta_\rho)^{-1}\cdot H\cdot(\Delta_\rho)^{-1},$$

then

$$G_H(\rho)^{-1}=(\Delta_\rho)\cdot K_H\cdot(\Delta_\rho).$$

Hence

$$\begin{aligned}\mathbf{g}^H(\rho)(\sigma,\sigma)=&\Big(\Psi,G_H(\rho)^{-1}G_H(\rho)G_H(\rho)^{-1}\Psi\Big)_{L^2}\\ =&\Big(\Psi,G_H(\rho)^{-1}\Psi\Big)_{L^2}\\ =&\Big(\Psi,\Delta_\rho\cdot K_H\cdot\Delta_\rho\Psi\Big)_{L^2}\\ =&\int\int\Psi(x)\nabla_x\cdot\Big(\rho(x)\nabla_xK_H(x,y,\rho)\nabla_y\cdot(\rho(y)\nabla_y\Psi(y))\Big)dxdy\\ =&-\int\int\nabla_x\Psi(x)\rho(x)\nabla_xK_H(x,y,\rho)\nabla_y\cdot(\rho(y)\nabla_y\Psi(y))dxdy\\ =&\int\int\nabla_x\nabla_yK_H(x,y,\rho)(\nabla_x\Psi(x),\nabla_y\Psi(y))\rho(x)\rho(y)dxdy,\end{aligned}$$

where the last two equalities apply the integration by parts w.r.t. variables x, y, respectively. □

We are now ready to formulate transport Newton's flows in term of Stein variational gradient flows.

Proposition 10 (Stein–transport Newton's flows)**.** *The transport Newton's flow of energy $\mathcal{E}(\rho)$ satisfies*

$$\partial_t\rho(t,x) = \nabla_x \cdot \Big(\rho(t,x)\int \rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho), \nabla_y\delta\mathcal{E}(\rho)(y))dy\Big).$$

In particular, consider $\mathcal{E}(\rho)$ as the KL divergence functional, i.e. $\mathcal{E}(\rho) = \int \rho\log\frac{\rho}{e^{-f}}dx$, then the transport Newton's flow satisfies

$$\begin{aligned}\partial_t\rho = &\quad \nabla_x \cdot \Big(\rho(t,x)\int \rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho), \nabla_y f(y)dy\Big)\\ &- \nabla_x \cdot \Big(\rho(t,x)\int \rho(t,y)\nabla_x\Delta_y K_H(x,y,\rho)dy\Big).\end{aligned} \tag{15}$$

Proof. From Theorem 4, the transport Hessian gradient flow satisfies the following equation

$$\begin{cases}\partial_t\rho - \nabla\cdot(\rho\nabla\Phi^\xi) = 0\\ -\displaystyle\int H(x,y,\rho)\Phi^\xi(y,\rho)dy = \nabla\cdot(\rho\nabla\delta\mathcal{E}(\rho)).\end{cases} \tag{16}$$

Denote $\Phi = -\Phi^\xi$. We notice that the second equation of (16) satisfies

$$\begin{aligned}\Phi(x,\rho) =& \int K_H(x,y,\rho)\nabla_y\cdot(\rho(y)\nabla_y\delta\mathcal{E}(\rho))dy\\ =& -\int(\nabla_y K_H(x,y,\rho), \nabla_y\delta\mathcal{E}(\rho))\rho(y)dy.\end{aligned}$$

Hence the first equation of (16) satisfies

$$\begin{aligned}\partial_t\rho =& -\nabla\cdot(\rho\nabla\Phi)\\ =&\nabla_x\cdot\Big(\rho\nabla_x\int(\nabla_y K_H(x,y,\rho), \nabla_y\delta\mathcal{E}(\rho)(y))\rho(t,y)dy\Big)\\ =&\nabla_x\cdot\Big(\rho\int\rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho), \nabla_y\delta\mathcal{E}(\rho)(y))dy\Big),\end{aligned}$$

which finishes the proof of first part. Denote $\mathcal{E}(\rho)$ as the KL divergence. Then $\delta\mathcal{E}(\rho) = \log\frac{\rho}{e^{-f}} + 1$. Hence

$$\begin{aligned}
\partial_t\rho = & \nabla_x\cdot\Big(\rho(t,x)\int\rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho),\nabla_y\log\frac{\rho(t,y)}{e^{-f(y)}}dy\Big)\\
= & \nabla_x\cdot\Big(\rho(t,x)\int\rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho),\nabla_y\log\rho(t,y))dy\Big)\\
& + \nabla_x\cdot\Big(\rho(t,x)\int\rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho),\nabla_y f(y))dy\Big)\\
= & \nabla_x\cdot\Big(\rho(t,x)\int(\nabla_x\nabla_y K_H(x,y,\rho),\nabla_y\rho(t,y))dy\Big)\\
& + \nabla_x\cdot\Big(\rho(t,x)\int\rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho),\nabla_y f(y))dy\Big)\\
= & -\nabla_x\cdot\Big(\rho(t,x)\int(\nabla_x\Delta_y K_H(x,y,\rho)\rho(t,y)dy\Big)\\
& + \nabla_x\cdot\Big(\rho(t,x)\int\rho(t,y)(\nabla_x\nabla_y K_H(x,y,\rho),\nabla_y f(y)dy\Big),
\end{aligned}$$

where the third equality we use the fact that $\rho\nabla\log\rho = \rho$, and the last equality holds by integration by parts for variable y. □

Proposition 11 (Particle formulation of Transport Newton's flow)**.** *Denote $X_t \sim \rho$, $Y_t \sim \rho$ are two independent identical processes, then the particle formulation of transport Newton's flow of energy $\mathcal{E}(\rho)$ satisfies*

$$\frac{d}{dt}X_t = -\mathbb{E}_{Y_t\sim\rho}\big[\nabla_x\nabla_y K_H(X_t,Y_t,\rho),\nabla_y\delta\mathcal{E}(\rho)(Y_t)\big].$$

In particular, if $\mathcal{E}(\rho)$ is the KL divergence functional, i.e. $\mathcal{E}(\rho) = \int\rho(x)\log\frac{\rho(x)}{e^{-f(x)}}dx$, where e^{-f} is a given target distribution with $f\in C^\infty(M)$, then the particle formulation of transport Newton's flow satisfies

$$\frac{d}{dt}X_t = -\mathbb{E}_{Y_t\sim\rho}\big[\nabla_x\nabla_y K_H(X_t,Y_t,\rho),\nabla_y f(Y_t)\big] + \mathbb{E}_{Y_t\sim\rho}\big[\nabla_x\Delta_y K_H(X_t,Y_t,\rho)\big].$$

Proof. Here the particle flow follows directly from the definition of Kolmogorov forward operator. □

Remark 8 (Transport Newton kernels)*.* Hence for computing the transport Newton's flow, the major issue is to approximate the inverse of kernel function defined in (11). For example, if $\mathcal{E}(\rho) = \mathrm{D}_{\mathrm{KL}}(\rho\|e^{-f}) = \int\rho\log\frac{\rho}{e^{-f}}dx$, then we have to consider

$$K_H(x,y,\rho) = \Big(\nabla^2\colon(\rho\nabla^2) - \nabla\cdot(\rho\nabla^2 f\nabla)\Big)^{-1}(x,y).$$

For this choice of kernel function, the asymptotic convergence rate of gradient flow (15) is of Newton type. Here the challenge is that we need to approximate the inverse of second order Laplacian operator by a kernel function K_H. This selection provides us hints for designing mean field kernel functions for accelerating Stein variational gradient; see related studies in [14, 15].

Remark 9 (Connecting Stein metric and transport metric). It is also worth mentioning that if $\mathcal{E}(\rho) = \frac{1}{2}\int x^2\rho(x)dx$, then the transport Hessian metric is the L^2 Wasserstein metric. From the equivalent relation between Stein metric and transport Hessian metric, we notice that L^2–Wasserstein metric is a particular mean field version of Stein metric. In details, if we choose the kernel function by

$$K_H(x,y,\rho) = \Big(-\nabla\cdot(\rho\nabla)\Big)^{-1}(x,y),$$

then $\mathbf{g}^H$ forms the L^2–Wasserstein metric $\mathbf{g}$. Hence, the corresponding Stein variational derivative is exactly the L^2-Wasserstein gradient.

Approximating kernel functions K_H in above remarks are interesting future directions.

5. Transport Hessian Hamiltonian flows

In this section, we present several dynamics associated with transport Hessian Hamiltonian flows.

To derive equations, we first consider the variational formulation (action functional) for the proposed Hamiltonian flow. Consider

$$\inf_{\rho}\Big\{\int_0^1 \frac{1}{2}\mathbf{g}^H(\partial_t\rho,\partial_t\rho) - \mathcal{F}(\rho)dt\Big\},$$

where the infimum is taken among all possible density path with suitable boundary conditions on initial and terminal densities ρ^0, ρ^1 respectively. Denote $\partial_t\rho = \Delta_\rho\Phi$, then the above action functional can be reformulated as follows:

$$\begin{aligned}
&\inf_{\rho,\Phi}\Big\{\int_0^1 \frac{1}{2}[\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi,\Phi)] - \mathcal{F}(\rho)dt\Big\}\\
=&\inf_{\rho,\Phi}\Big\{\int_0^1 \frac{1}{2}\Big[\int\int \nabla^2_{xy}\delta^2\mathcal{E}(\rho)(x,y)\nabla_x\Phi(t,x)\nabla\Phi(t,y)\rho(t,x)\rho(t,y)dxdy\\
&\qquad + \int \nabla^2_x\delta\mathcal{E}(\rho)(x)(\nabla\Phi(t,x),\nabla\Phi(t,x))\rho(t,x)dx\Big] - \mathcal{F}(\rho)dt\Big\}
\end{aligned}\tag{17}$$

where the infimum is taken among all possible density and potential functions $(\rho,\Phi)\colon [0,1]\times M] \to \mathbb{R}^2$, such that the continuity equation holds with the gradient vector drift function:

$$\partial_t\rho(t,x) + \nabla\cdot(\rho(t,x)\nabla\Phi(t,x)) = 0.$$

Here suitable boundary conditions are given on initial and terminal densities.

We are ready to derive Hamiltonian flows in Hessian density manifold.

Theorem 12 (Transport Hessian Hamiltonian flows)**.** *The Hamiltonian flow in* $(\mathcal{P}, \mathbf{g}^H(\rho))$ *satisfies*

$$\begin{cases} \partial_t\rho + \nabla\cdot(\rho\nabla\Phi) = 0 \\ \partial_t\Psi + (\nabla\Phi, \nabla\Psi) + \delta\mathcal{F}(\rho) = \frac{1}{2}\int \nabla^2_{yz}\delta^3\mathcal{E}(\rho)(x,y,z)(\nabla_y\Phi(t,y), \nabla_z\Phi(t,z))\rho(t,y)\rho(t,z)dydz \\ \qquad\qquad + \int \nabla^2_{xy}\delta^2\mathcal{E}(\rho)(x,y)(\nabla_x\Phi(t,x), \nabla_y\Phi(t,y))\rho(t,y)dy \\ \qquad\qquad + \frac{1}{2}\int \nabla^2_{y}\delta^2\mathcal{E}(\rho)(x,y)(\nabla_y\Phi(t,y), \nabla_y\Phi(t,y))\rho(t,y)dy \\ \qquad\qquad + \frac{1}{2}\nabla^2_x\delta\mathcal{E}(\rho)(x)(\nabla_x\Phi(t,x), \nabla_x\Phi(t,x)), \end{cases} \tag{18}$$

where Φ*,* Ψ *satisfy the Poisson equation*

$$\nabla_x\cdot(\rho(t,x)\nabla_x\Psi(t,x)) = \nabla_x\cdot\Big(\rho(t,x)[\nabla_x\int \nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla_y\Phi(t,y)\rho(t,y)dy + \nabla^2_x\delta\mathcal{E}(\rho)\nabla\Phi]\Big). \tag{19}$$

Here, δ^3 *denotes the* L^2 *third variational derivative.*

Proof. We apply the Lagrange multiplier method to solve variational problem (21). Denote $\Psi\colon [0,1]\times M\to\mathbb{R}$ as the Lagrange multiplier, then the Lagrangian functional in density space forms

$$\mathcal{L}(\rho,\Phi,\Psi) = \int_0^1 \frac{1}{2}\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi,\Phi) - \mathcal{F}(\rho)dt + \int_0^1\int \Psi\Big(\partial_t\rho + \nabla\cdot(\rho\nabla\Phi)\Big)dxdt.$$

Hence $\delta_\rho\mathcal{L}=0$, $\delta_\Phi\mathcal{L}=0$, $\delta_\Psi\mathcal{L}=0$ imply the fact that

$$\begin{cases} \frac{1}{2}\delta_\rho\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi,\Phi) - \delta\mathcal{F}(\rho) - \partial_t\Psi - (\nabla\Phi,\nabla\Psi) = 0 \\ \nabla_x\cdot(\rho(t,x)\nabla_x\int \nabla_y\delta^2\mathcal{E}(\rho)(x,y)\nabla_y\Phi(t,y)\rho(t,y)dy) + \nabla_x\cdot(\rho\nabla^2_x\delta\mathcal{E}(\rho)\nabla_x\Phi) = \nabla\cdot(\rho\nabla\Psi) \\ \partial_t\rho + \nabla\cdot(\rho\nabla\Phi) = 0, \end{cases}$$

which finishes the proof. □

In above derivations, we notice that Ψ is the momentum variable in $\mathbf{g^H}$ and Φ is the momentum variable in $\mathbf{g}$. And the Hamiltonian in $(\mathcal{P}, g^{\mathbf{H}})$ satisfies

$$\mathcal{H}(\rho,\Psi) = \frac{1}{2}\mathrm{Hess}^*_{\mathbf{g}}\mathcal{E}(\rho)(\Phi,\Phi) + \mathcal{F}(\rho),$$

where Φ,Ψ are associated with the Poisson equation (19). We notice the Hamiltonian flow (18) is equivalent to the following formulation

$$\partial_{tt}\rho + \mathbf{\Gamma^H}(\rho)(\partial_{\mathbf{t}}\rho, \partial_{\mathbf{t}}\rho) = -\mathrm{grad}_{\mathbf{g^H}}\mathcal{F}(\rho).$$

where $\Gamma^H(\rho)$ is the Christoffel symbol in transport Hessian metric. We omit the detailed formulations fo Γ^H; see details in [24]. In particular, when $\mathcal{F}(\rho)=0$, the above Hamiltonian flow forms the geodesic equation in $(\mathcal{P}, \mathbf{g}^H)$.

5.1. **Examples.** We next present several examples of transport Hessian Hamiltonian flows.

Example 10 (Linear energy). *Consider $\mathcal{E}(\rho)=\int E(x)\rho(x)dx$. Then the transport Hessian Hamiltonian flow forms*

$$\begin{cases}\partial_t\rho+\nabla\cdot(\rho\nabla\Phi)=0\\ \partial_t\Psi+(\nabla\Phi,\nabla\Psi)-\frac{1}{2}\nabla^2E(x)(\nabla\Phi,\nabla\Phi)+\delta\mathcal{F}(\rho)=0\\ \nabla\cdot(\rho\nabla\Psi)=\nabla\cdot(\rho\nabla^2E(x)\nabla\Phi).\end{cases}$$

Again, if $E(x)=\frac{1}{2}\int x^2\rho(x)dx$, then the Hamiltonian flow satisfies

$$\begin{cases}\partial_t\rho+\nabla\cdot(\rho\nabla\Phi)=0\\ \partial_t\Phi+\frac{1}{2}(\nabla\Phi,\nabla\Phi)+\delta\mathcal{F}(\rho)=0.\end{cases}$$

The above equation system is known as the Wasserstein/transport Hamiltonian flow [10, 24]*. It is a particular format for compressible Euler equations. In particular, if $\mathcal{F}(\rho)=0$, it is the Wasserstein/transport geodesics equation* [40]*.*

Example 11 (Interaction energy). *Consider $\mathcal{E}(\rho)=\frac{1}{2}\int\int W(x,y)\rho(x)\rho(y)dxdy$. Then the transport Hessian Hamiltonian flow satisfies*

$$\begin{cases}\partial_t\rho+\nabla\cdot(\rho\nabla\Phi)=0\\ \partial_t\Psi+(\nabla\Phi,\nabla\Psi)+\delta\mathcal{F}(\rho)=\int\nabla^2_{xy}W(x,y)(\nabla\Phi(t,x),\nabla\Phi(t,y))\rho(t,y)dy\\ \qquad\qquad+\frac{1}{2}\int\nabla^2_yW(x,y)(\nabla_y\Phi(t,y),\nabla_y\Phi(t,y))\rho(t,y)dy\\ \qquad\qquad+\frac{1}{2}\int\nabla^2_xW(x,y)(\nabla_x\Phi(t,x),\nabla_x\Phi(t,x))\rho(t,y)dy\\ \nabla\cdot(\rho\nabla\Psi)=\nabla_x\cdot\Big(\rho(t,x)\int[\nabla^2_{xy}W(x,y)\nabla_y\Phi(t,y)+\nabla^2_xW(x,y)\nabla\Phi(t,x)]\rho(t,y)dy\Big).\end{cases}$$

Example 12 (Entropy). *Consider $\mathcal{E}(\rho)=\int f(\rho)(x)dx$. Then the transport Hessian Hamiltonian flow satisfies*

$$\begin{cases}\partial_t\rho+\nabla\cdot(\rho\nabla\Phi)=0\\ \partial_t\Psi+(\nabla\Phi,\nabla\Psi)-\frac{1}{2}\|\nabla^2\Phi\|^2p'(\rho)-\frac{1}{2}|\Delta\Phi|^2p_2'(\rho)+\delta\mathcal{F}(\rho)=0\\ -\nabla\cdot(\rho\nabla\Psi)=\nabla^2\colon(p(\rho)\nabla^2\Phi)+\Delta(p_2(\rho)\Delta\Phi).\end{cases}$$

We next present several examples for transport Hessian flows of entropies.

Example 13 (Quadratic entropy). *Consider $\mathcal{E}(\rho)=\frac{1}{2}\int\rho(x)^2dx$. Then the transport Hessian Hamiltonian flow satisfies*

$$\begin{cases}\partial_t\rho+\nabla\cdot(\rho\nabla\Phi)=0\\ \partial_t\Psi+(\nabla\Phi,\nabla\Psi)-\frac{1}{2}\|\nabla^2\Phi\|^2\rho-\frac{1}{2}|\Delta\Phi|^2\rho+\delta\mathcal{F}(\rho)=0\\ -\nabla\cdot(\rho\nabla\Psi)=\frac{1}{2}\nabla^2\colon(\rho\nabla^2\Phi)+\frac{1}{2}\Delta(\rho\Delta\Phi).\end{cases}$$

Example 14 (Negative Boltzmann–Shannon entropy)**.** *Consider* $\mathcal{E}(\rho) = \int \rho(x) \log \rho(x) dx$. *Then the transport Hessian Hamiltonian flow satisfies*

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \Phi) = 0 \\ \partial_t \Psi + (\nabla \Phi, \nabla \Psi) - \frac{1}{2}\|\nabla^2 \Phi\|^2 + \delta \mathcal{F}(\rho) = 0 \\ -\nabla \cdot (\rho \nabla \Psi) = \nabla^2 \colon (\rho \nabla^2 \Phi). \end{cases} \tag{20}$$

5.2. **Connections with Shallow Water's equation.** We next demonstrate that equation (20) connects with the Shallow water equation. If $M = \mathbb{T}^1$ and

$$\mathcal{E}(\rho) = \int \rho \log \rho dx, \qquad \mathcal{F}(\rho) = -\frac{1}{2} \int \rho^2 dx,$$

then equation (20) is known as the Shallow water equation in fluid dynamics; see [4] and many references therein. In other words, in one dimensional sample space, denote $v(t, x) = \nabla \Phi(t, x)$, then equation (20) forms

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho v) = 0 \\ \partial_t \Psi + (v, \nabla \Psi) - \frac{1}{2}\|\nabla v\|^2 - \rho = 0 \\ -\rho \nabla \Psi = \nabla \cdot (\rho \nabla v). \end{cases}$$

The above equation system is the minimizer of variational problem

$$\inf_{\rho, v = \nabla \Phi} \Big\{ \int_0^1 \int \Big(\|\nabla v(t,x)\|^2 \rho(t,x) + \frac{1}{2}\rho(t,x)^2 \Big) dx dt \colon \partial_t \rho(t,x) + \nabla \cdot (\rho(t,x) v(t,x)) = 0 \Big\}, \tag{21}$$

where the infimum is taken among continuity equations with **gradient vector fields** and suitable boundary conditions on initial and terminal densities. Here Ψ is the Lagrange multiplier of the continuity equation.

Remark 10. We remark that the Hamiltonian flows in Hessian density manifold of negative Boltzmann–Shanon entropy coincides with the ones using Hessian operators in diffeomorphism space [19] or semi-invariant metric [4] in one dimensional sample space. Other than one dimensional space, these metrics or related variational structures are different. In other words, consider the variational problem

$$\inf_{\rho, v} \Big\{ \int_0^1 \int \|\nabla v(t,x)\|^2 \rho(t,x) + \frac{\rho^2}{2} dx \colon \partial_t \rho(t,x) + \nabla \cdot (\rho(t,x) v(t,x)) = 0 \Big\}, \tag{22}$$

where the infimum is taken among continuity equations with all vector fields and suitable initial and terminal densities. When $\dim(M) = 1$, variational problems (21) and (22) are equivalent. If $\dim(M) \neq 1$, they are different formulations.

To summarize, we provide the variational formulation of Shallow Water equation based on both negative Boltzmann–Shannon entropy and transport Hessian metrics.

6. Finite dimensional Transport Hessian metric

In this section, we demonstrate the transport Hessian metric in finite dimensional probability models. In other words, we pull back the transport Hessian metric into a finite dimensional parameter space. Several examples are provided, including one dimensional probability models and Gaussian families.

For the simplicity of presentation, we focus on the transport Hessian metric of negative Shannon–Boltzmann entropy defined in (6). Consider a statistical model defined by a triplet (Θ, M, ρ). For the simplicity of presentation, we assume $\Theta \subset \mathbb{R}^d$ and $\rho\colon \Theta \to \mathcal{P}(M)$ is a parametrization function. In this case, $\rho(\Theta) \subset \mathcal{P}(M)$. We assume that the parameterization map ρ is locally injective and smooth. Given $\rho(\theta, \cdot) \in \rho(\Theta)$, denote the tangent space of probability density space by

$$T_{\rho_\theta}\rho(\Theta) = \big\{\dot\rho_\theta = (\nabla_\theta \rho_\theta, \dot\theta) \in C^\infty(M)\colon \dot\theta \in T_\theta\Theta\big\}.$$

We define a Riemannian metric $G_H(\theta) = (G_H(\theta)_{ij})_{1\leq i,j\leq d} \in \mathbb{R}^{d\times d}$ on Θ as the pull-back of $\mathbf{g}^H$ on $\mathcal{P}$, i.e.

$$\dot\theta^{\mathsf{T}} G_H(\theta)\dot\theta = \mathbf{g}^H_{\rho_\theta}(\dot\rho_\theta, \dot\rho_\theta). \tag{23}$$

Here G_H is a semi-positive matrix function and (Θ, G_H) is named the statistical manifold. Denote ∇ as the Euclidean derivative w.r.t x and ∂_θ as the Euclidean derivative w.r.t. θ. We present the transport Hessian metric as follows.

Definition 13 (Transport Hessian information matrix)**.** *The transport Hessian metric in statistical manifold* (Θ, G_H) *satisfies*

$$G_H(\theta)_{ij} = \int tr\big(\nabla^2\Phi_i(x;\theta), \nabla^2\Phi_j(x;\theta)\big)\rho(x;\theta)dx,$$

where

$$-\nabla\cdot(\rho(x;\theta)\nabla\Phi_k(x;\theta)) = \nabla_{\theta_k}\rho(x;\theta), \qquad k = i, j.$$

Remark 11. Following studies in [30], we call this finite dimensional metric the *transport Hessian information matrix*. This is the other generalization of information matrix studied in [1, 2, 3]. We leave the systematic studies of transport information matrix in both statistics and machine learning inference problems for the future work.

Proof. The definition follows from the definition of pull-back operator. Notice that

$$\begin{aligned} G_H(\theta)_{ij} =& \mathbf{g}^H(\rho)(\nabla_{\theta_i}\rho, \nabla_{\theta_j}\rho) \\ =& \int \Big(\nabla_{\theta_i}\rho(x;\theta), (\nabla\cdot\rho\nabla)^{-1}\Big(\nabla^2\colon \rho\nabla^2\Big)(\nabla\cdot\rho\nabla)^{-1}\nabla_{\theta_j}\rho(x;\theta)\Big)dx. \end{aligned}$$

By denoting $-\nabla\cdot(\rho\nabla\Phi_k) = \nabla_\theta\rho$, we finish the proof. □

We next present several closed formulas for finite dimensional transport Hessian information metrics.

Example 15 (One dimensional sample space)**.** *If* $M = \mathbb{T}^1$*, then*

$$\nabla\Phi_k(x;\theta) = -\frac{1}{\rho(x;\theta)}\partial_{\theta_k}F(x;\theta),$$

where $k = 1, \cdots, d$ and F is the cumulative distribution function of $\rho(x;\theta)$ with $F(x;\theta) = \int_0^x \rho(y;\theta)dy$. In this case, we have

$$
\begin{aligned}
G_H(\theta)_{ij} =& \int \nabla^2\Phi_i \nabla^2\Phi_j \rho dx\\
=& \int \nabla(\frac{\partial_{\theta_i}F}{\rho})\cdot\nabla(\frac{\partial_{\theta_j}F}{\rho})\rho dx\\
=& \int (\frac{\nabla\partial_{\theta_i}F}{\rho} + \partial_{\theta_i}F\nabla\frac{1}{\rho})(\frac{\nabla\partial_{\theta_j}F}{\rho} + \partial_{\theta_j}F\nabla\frac{1}{\rho})\rho dx\\
=& \int (\frac{\partial_{\theta_i}\rho}{\rho} + \partial_{\theta_i}F\nabla\frac{1}{\rho})(\frac{\partial_{\theta_j}\rho}{\rho} + \partial_{\theta_j}F\nabla\frac{1}{\rho})\rho dx\\
=& \int \frac{\partial_{\theta_i}\rho\partial_{\theta_j}\rho}{\rho} + \partial_{\theta_i}F\nabla\frac{1}{\rho}\partial_{\theta_j}\rho + \partial_{\theta_j}F\nabla\frac{1}{\rho}\partial_{\theta_i}\rho + \partial_{\theta_i}F\partial_{\theta_j}F(\nabla\frac{1}{\rho})^2\rho dx\\
=& \quad \int \big(\rho\partial_{\theta_i}\log\rho\partial_{\theta_j}\log\rho\big)dx\\
&+\int\big(-\partial_{\theta_i}F\nabla\log\rho\partial_{\theta_j}\log\rho - \partial_{\theta_j}F\nabla\log\rho\partial_{\theta_i}\log\rho + \frac{\partial_{\theta_i}F\partial_{\theta_j}F}{\rho}(\nabla\log\rho)^2\big)dx,
\end{aligned}
$$

where we use the fact that $\nabla\frac{1}{\rho} = -\frac{\nabla\rho}{\rho^2} = -\frac{1}{\rho}\nabla\log\rho$, $\nabla\log\rho = \frac{\nabla\rho}{\rho}$ and $\partial_\theta\log\rho = \frac{\partial_\theta\rho}{\rho}$ in the last equality. We notice that the transport Hessian metric is a modification of Fisher–Rao metric with transport Levi-Civita connection [24].

Remark 12. We compare all information matrices in one dimensional sample space as follows:

$$G_F(\theta)_{ij} = \int \frac{\partial_{\theta_i}\rho(x;\theta)\partial_{\theta_j}\rho(x;\theta)}{\rho(x;\theta)}dx \qquad \text{Fisher information matrix}$$

$$G_W(\theta)_{ij} = \int \frac{\partial_{\theta_i}F(x;\theta)\partial_{\theta_j}F(x;\theta)}{\rho(x;\theta)}dx \qquad \text{Wasserstein information matrix}$$

$$G_H(\theta)_{ij} = \int \nabla(\frac{\partial_{\theta_i}F(x;\theta)}{\rho(x;\theta)})\cdot\nabla(\frac{\partial_{\theta_j}F(x;\theta)}{\rho(x;\theta)})\rho(x;\theta)dx \quad \text{Transport Hessian information matrix}$$

Example 16 (Gaussian family)**.** *Consider a Gaussian family by*

$$p(x;\theta) = \frac{1}{\sqrt{2\pi\Sigma}}e^{-\frac{1}{2}(x-m)^\mathsf{T}\Sigma^{-1}(x-m)}, \tag{24}$$

where $x \in \mathbb{R}^d$, $\theta = (m, \Sigma)$, $m \in \mathbb{R}^d$ is a mean value vector, and $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite covariance matrix. In this case, the continuity equation with gradient drift has a closed form solution. Denote $\dot\theta = (\dot m, \dot\Sigma) \in T_\theta\Theta$. We can check that if

$$-\nabla\cdot(\rho\nabla\Phi_k) = (\nabla_\theta\rho, \dot\theta),$$

then there exists a symmetric matrix $S \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$, such that

$$\Phi(x) = \frac{1}{2}x^\mathsf{T}Sx + b^\mathsf{T}x,$$

and

$$\dot\Sigma = \Sigma S + S\Sigma, \qquad \dot m = b.$$

Hence the transport Hessian metric in Gaussian family satisfies

$$G_H(\Sigma)((\dot m_1,\dot\Sigma_1),(\dot m_2,\dot\Sigma_2)) = Tr(S_1S_2),$$

where

$$\dot\Sigma_1=\Sigma S_1+S_1\Sigma,\quad \dot\Sigma_2=\Sigma S_2+S_2\Sigma.$$

We notice that the transport Hessian metric is degenerate in the direction of mean values. Besides, we consider a Gaussian family with zero mean in one dimensional space. In this case,

$$G_H(\Sigma)(\dot\Sigma_1,\dot\Sigma_2)=S_1S_2=\frac{\dot\Sigma_1\dot\Sigma_2}{4\Sigma^2}.$$

Hence the distance function has a closed form solution. Notice

$$\begin{aligned}
Dist_H(\Sigma_0,\Sigma_1)^2 &= \inf_{\Sigma\colon[0,1]\to\mathbb{R}_+}\Big\{\int_0^1 G_H(\frac{d}{dt}\Sigma(t),\frac{d}{dt}\Sigma(t))dt\colon \Sigma_0,\ \Sigma_1\ \textit{fixed}\Big\}\\
&= \inf_{\Sigma\colon[0,1]\to\mathbb{R}_+}\Big\{\int_0^1 \frac{(\frac{d}{dt}\Sigma(t))^2}{4\Sigma^2}dt\colon \Sigma_0,\ \Sigma_1\ \textit{fixed}\Big\}\\
&= \inf_{\Sigma\colon[0,1]\to\mathbb{R}_+}\Big\{\int_0^1 \frac{1}{4}(\frac{d}{dt}\log\Sigma(t))^2dt\ :\ \Sigma_0,\ \Sigma_1\ \textit{fixed}\Big\},
\end{aligned}$$

where we use the fact that $\frac{d}{dt}\log\Sigma(t)=\frac{\frac{d}{dt}\Sigma(t)}{\Sigma(t)}$. *Hence the geodesics satisfies* $\frac{d^2}{dt^2}\log\Sigma(t)=0$, *and* $\log\Sigma(t)=(1-t)\log\Sigma_0+t\log\Sigma_1$. *Hence we derive the closed form solution for the transport Hessian metric:*

$$Dist_H(\Sigma_0,\Sigma_1)=\frac{1}{2}\|\log\Sigma_0-\log\Sigma_1\|.$$

In this case, we observe that the transport Hessian metric coincides with the Fisher–Rao metric for covariance matrix in Gaussian families.

Example 17 (Generalized transport Hessian metric for Gaussian family)**.** *In this example, we consider a finite dimensional transport Hessian metric for energy* $\mathcal{E}(\rho)=\int(\rho\log\rho+\frac{x^2}{2}\rho)dx$. *In this case, the metric* $\mathbf{g}^H$ *forms*

$$\mathbf{g}^H(\rho)(\sigma_1,\sigma_2)=\int\Big\{tr(\nabla^2\Phi_1(x),\nabla^2\Phi_2(x))+(\nabla\Phi_1(x),\nabla\Phi_2(x))\Big\}\rho(x)dx.$$

where $\sigma_i=-\nabla\cdot(\rho\nabla\Phi_i)$, $i=1,2$. *Similarly as in Example 16, we consider the Gaussian family* (24) *in one dimensional sample space. Then*

$$G_H(m,\Sigma)((\dot m_1,\dot\Sigma_1),(\dot m_2,\dot\Sigma_2))=S_1S_2+S_1\Sigma S_2+b_1b_2=\frac{\dot\Sigma_1\dot\Sigma_2}{4\Sigma^2}+\frac{\dot\Sigma_1\dot\Sigma_2}{4\Sigma}+\dot m_1\dot m_2.$$

In this case, the transport Hessian distance function also has a closed form solution. Similarly,

$$
\begin{aligned}
&Dist_H((m_0,\Sigma_0),(m_1,\Sigma_1))^2\\
=&\inf_{\Sigma\colon[0,1]\to\mathbb{R}_+}\Big\{\int_0^1 G_H(\frac{d}{dt}\Sigma(t),\frac{d}{dt}\Sigma(t))dt\colon (m_0,\Sigma_0),\ (m_1,\Sigma_1)\ \textit{fixed}\Big\}\\
=&\inf_{\Sigma\colon[0,1]\to\mathbb{R}_+}\Big\{\int_0^1 (\frac{d}{dt}\Sigma(t))^2(\frac{1}{4\Sigma(t)^2}+\frac{1}{4\Sigma(t)})+(\frac{d}{dt}m(t))^2dt\colon (m_0,\Sigma_0),\ (m_1,\Sigma_1)\ \textit{fixed}\Big\}\\
=&\inf_{\Sigma\colon[0,1]\to\mathbb{R}_+}\Big\{\int_0^1 \|\frac{d}{dt}\big(\sqrt{\Sigma(t)+1}-tanh^{-1}(\sqrt{\Sigma(t)+1})\big)\|^2+\|\frac{d}{dt}m(t)\|^2dt\ :\ (m_0,\Sigma_0),\ (m_1,\Sigma_1)\ \textit{fixed}\Big\},
\end{aligned}
$$

where we use the fact that $\frac{d}{dt}\big(\sqrt{\Sigma(t)+1}-tanh^{-1}(\sqrt{\Sigma(t)+1})\big)=\frac{1}{2}\sqrt{\frac{1}{\Sigma(t)^2}+\frac{1}{\Sigma(t)}}\frac{d}{dt}\Sigma(t)$. *Hence the geodesics forms*

$$
\begin{cases}
\dfrac{d^2}{dt^2}\big(\sqrt{\Sigma(t)+1}-tanh^{-1}(\sqrt{\Sigma(t)+1})\big)=0\\
\dfrac{d^2}{dt^2}m(t)=0.
\end{cases}
$$

Hence the geodesics satisfies

$$
\begin{cases}
\sqrt{\Sigma(t)+1}-tanh^{-1}(\sqrt{\Sigma(t)+1})=(1-t)(\sqrt{\Sigma_0+1}-tanh^{-1}(\sqrt{\Sigma_0+1}))\\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad+t(\sqrt{\Sigma_1+1}-tanh^{-1}(\sqrt{\Sigma_1+1}))\\
m(t)=(1-t)m_0+tm_1
\end{cases}
$$

And we derive the closed form solution for the transport Hessian metric:

$$
\begin{aligned}
&Dist_H((m_0,\Sigma_0),(m_1,\Sigma_1))^2\\
=&\|m_0-m_1\|^2+\|\sqrt{\Sigma_0+1}-\sqrt{\Sigma_1+1}-(tanh^{-1}(\sqrt{\Sigma_0+1})-tanh^{-1}(\sqrt{\Sigma_1+1}))\|.
\end{aligned}
$$

There are many other interesting closed form solutions for transport Hessian metrics. We shall derive them in the future work.

7. Discussions

In this paper, we study the Hessian metric of given energy for optimal transport metrics. We name the density space with transport Hessian metric the Hessian density manifold. We discover several connections between dynamics in Hessian density manifold and math physics equations, including Shallow water's equations and heat equations. In particular, we demonstrate that a transport Hessian metric induces a mean-field kernel for Stein metric, following which the Stein variational derivative forms the transport Newton's direction.

Following the transport Hessian metric, there are several future directions among Hessian metric, scientific computing methods, and machine learning sampling algorithms. Firstly, we notice that the optimal transport metrics belong to a particular class of Hessian metric in density manifold. For these transport Hessian metrics, are there any other formulations, such as Monge problems, Monge-Amperé equations or Kantorovich dualities as in optimal transport? Secondly, can we formulate *transport divergence functionals* based on transport Hessian metrics? We will conduct this direction following information geometry; Thirdly, the proposed structure can involve several variational formulations for math physics equations. E.g. can this method be useful for designing numerical algorithms towards Shallow water's equations? Lastly and most importantly, the proposed metric provides a hint for selecting the kernel to accelerate Stein variational derivative. We shall propose numerical methods to approximate kernels in the proposed Stein–transport Newton's flow. This could be useful in designing mean-field Markov chain Monte Carlo methods.

References

[1] S. Amari. *Information Geometry and Its Applications.* Springer Publishing Company, Incorporated, 1st edition, 2016.

[2] S. Amari. Wasserstein statistics in 1D location-scale model. *arXiv:2003.05479*, 2020.

[3] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. *Information geometry*, volume 64. Springer, Cham, 2017.

[4] M. Bauer, and K. Modin. Semi-invariant Riemannian metrics in hydrodynamics. *Calculus of Variations and Partial Differential Equations volume* 59, 2020.

[5] D. Bakry and M. Émery. Diffusions hypercontractives. *Séminaire de probabilités de Strasbourg*, 19:177–206, 1985.

[6] E. Carlen. Conservative Diffusions. *Communications in math physics*, 94(3):293–315, 1984.

[7] Y. Chen and W. Li. Wasserstein natural gradient in statistical manifolds with continuous sample space. *arXiv:1805.08380*, 2018.

[8] S. Cheng and S. T. Yau. The real Monge-Ampére equation and affine flat structures *Proc. 1980 Beijing Symp. Differ. Geom. and Diff. Eqns.*, Vol. 1, pp. 339-370, 1982.

[9] S. Chow, W. Li, and H. Zhou. A discrete Schrödinger equation via optimal transport on graphs. *Journal of Functional Analysis*, 276(8):2440–2469, 2019.

[10] S. Chow, W. Li, and H. Zhou. Wasserstein Hamiltonian flow. *Journal of differential equation*, 268(3):1205–1219, 2020.

[11] S. Chow, W. Li, and H. Zhou. Entropy dissipation of Fokker-Planck equations on graphs. *Discrete and Continuous Dynamical Systems–A*, v.38, 2018.

[12] G. Conforti and M. Pavon. Extremal flows on Wasserstein space. *Extremal flows in Wasserstein space*, Journal of math physics, 59, 2018.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications. Wiley, New York, 1991.

[14] G. Detommaso, T. Cui, A. Spantini, Y. Marzouk, and R. Scheichl. A Stein variational Newton method. *NIPS*, 2018.

[15] A. Duncan, N. Nuesken, and L. Szpruch. On the geometry of Stein variational gradient descent. *arXiv:1912.00894*, 2019.

[16] Q. Feng, and W. Li. Generalized Gamma z calculus via sub-Riemannian density manifold. *arXiv:1910.07480*, 2019.

[17] A. Garbuno-Inigo, F. Hoffmann, W. Li , and A. Stuart. Interacting Langevin Diffusions: Gradient Structure And Ensemble Kalman Sampler. *SIAM Journal on Applied Dynamical Systems*, 19(1), 412–441, 2020.

[18] U. Gianazza, G. Savaré, and G. Toscani. The Wasserstein Gradient Flow of the Fisher Information and the Quantum Drift-diffusion Equation. *Archive for Rational Mechanics and Analysis*, volume 194, pages 133–220, 2009.

[19] D. Ionescu-Kruse. Variational derivation of the Green–Naghdi–Shallow–Water equations. *Journal of Nonlinear math physics*, 19, 1-12, 2012.

[20] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[21] B. Khesin, G. Misiolek, and K. Modin. Geometric Hydrodynamics of Compressible Fluids *arXiv:2001.01143*, 2020.

[22] J. D. Lafferty. The density manifold and configuration space quantization. *Transactions of the American Mathematical Society*, 305(2):699–741, 1988.

[23] F. Leger, and W. Li. Hopf-Cole transformation via generalized Schrdinger bridge problem. *arXiv:1901.09051*, 2019.

[24] W. Li. Geometry of probability simplex via optimal transport. *arXiv:1803.06360 [math]*, 2018.

[25] W. Li. Diffusion Hypercontractivity via Generalized Density Manifold. *arXiv:1907.12546*, 2019.

[26] W. Li, J. Lu and L. Wang. Fisher information regularization schemes for Wasserstein gradient flows. *arXiv:1907.02152*, 2019.

[27] W. Li and G. Montufar. Natural gradient via optimal transport. *Information Geometry*, 1, pages181–214, 2018.

[28] W. Li and G. Montufar. Ricci curvature for parametric statistics via optimal transport. *Information Geometry*, 2020.

[29] W. Li, and L. Ying. Transport Hessian gradient flows. *Research in the Mathematical Sciences*, 2019.

[30] W. Li, and J. Zhao. Wasserstein information matrix. *arXiv:1910.11248*, 2019.

[31] Q. Liu. Stein Variational Gradient Descent as Gradient Flow. *NIPS*, 2017.

[32] L. Malago, and G. Pistone. Natural Gradient Flow in the Mixture Geometry of a Discrete Exponential Family. *Entropy*, 17, 4215–4254, 2015.

[33] L. Malago, and G. Pistone. Combinatorial Optimization with Information Geometry: The Newton Method. *Entropy*, 16, 4260–4289, 2014.

[34] E. Nelson. Derivation of the Schrödinger Equation from Newtonian Mechanics. *Physical Review*, 150(4):1079–1085, 1966.

[35] E. Nelson. *Quantum Fluctuations.* Princeton series in physics. Princeton University Press, Princeton, N.J, 1985.

[36] Y. Wang, and W. Li. Accelerated information gradient flow. *arXiv:1909.02102*, 2019.

[37] Y. Wang, and W. Li. Information Newton's flow: second-order optimization method in probability space. *arXiv:2001.04341*, 2020.

[38] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.

[39] G. Perelman. The entropy formula for the Ricci flow and its geometric applications. *arXiv:math/0211159*, 2002.

[40] C. Villani. *Optimal Transport: Old and New.* Number 338 in Grundlehren Der Mathematischen Wissenschaften. Springer, Berlin, 2009.

E-mail address: wcli@ucla.math.edu

Mathematics department, UCLA, Los Angeles, CA 90095 U.S.A.