Calc. Var. (2023) 62:206
https://doi.org/10.1007/s00526-023-02550-2

Calculus of Variations

Generalized curvature for the optimal transport problem induced by a Tonelli Lagrangian

Yuchuan Yang[1]

Received: 27 April 2022 / Accepted: 25 July 2023 / Published online: 5 August 2023
© The Author(s) 2023

Abstract
We propose a generalized curvature that is motivated by the optimal transport problem on $\mathbb{R}^d$ with cost induced by a Tonelli Lagrangian L. We show that non-negativity of the generalized curvature implies displacement convexity of the generalized entropy functional on the L-Wasserstein space along C^2 displacement interpolants.

Mathematics Subject Classification 35A15 · 49Q22 · 53A99

1 Introduction

Given a Riemannian manifold (M, g), one may consider the optimal transport problem with cost given by squared Riemannian distance. This induces the *2-Wasserstein distance* W_2 on $\mathcal{P}_2(M)$, the space of probability measures on M with finite second moments (i.e. probability measures μ such that

$$\int_M d(x, x_0)^2 d\mu(x) < \infty$$

for every $x_0 \in M$). The metric space $(\mathcal{P}_2(M), W_2)$ is called the *2-Wasserstein space*, and is known to be a geodesic space ([21] Chapter 7).

In [17], Otto proposed that $\mathcal{P}_2(M)$ admits a formal Riemannian structure and developed a formal calculus on $\mathcal{P}_2(M)$. This later became what is known as *Otto calculus* [21] and was made rigorous by Ambrosio-Gigli-Savaré [2]. In particular, Otto calculus allows one to compute *displacement Hessians* of functionals along geodesics in $\mathcal{P}_2(M)$. This is useful for characterizing a displacement convex functional (i.e. convex along every geodesic) by the non-negativity of its displacement Hessian. In a seminal work by Otto and Villani [18], it was shown that the displacement convexity of the entropy functional is related to the Ricci curvature of (M, g). Since then, the notion of displacement convexity has been useful in

Communicated by Xavier Ros-Oton.

✉ Yuchuan Yang
yangyuchuan@ucla.edu

1 Department of mathematics, UCLA, California, USA

many other areas. For instance, it has inspired new heuristics and proofs of various functional inequalities [1, 8].

Further advances have been made towards understanding the relationships between the geometry of the underlying space and the induced geometry of $\mathcal{P}(M)$, the space of probability measures on M. In his Ph.D. thesis [19], Schachter studied the optimal transport problem on $\mathbb{R}^d$ with cost induced by a *Tonelli Lagrangian*. The case $d = 1$ was considered in [20], and this work was later used in [3] and [15].

In his work, Schachter developed an *Eulerian calculus*, extending the Otto calculus. Among the other contributions of his thesis, Schachter derived a canonical form for the displacement Hessians of functionals. Using Eulerian calculus, he found a new class of displacement convex functionals on S^1 [20], which includes those found by Carrillo and Slepčev in [7]. In the case when the cost is given by squared Riemannian distance, Schachter proved that his displacement Hessian agrees with Villani's displacement Hessian in [21], which is a quadratic form involving the *Bakry–Emery tensor*.

Summary of main results: In this manuscript, a generalized notion of curvature $\mathcal{K}_x$ (Definition 5.6) is proposed for the manifold $M = \mathbb{R}^d$ equipped with a general Tonelli Lagrangian L, and is given by

$$\mathcal{K}_x(\xi) := \operatorname{tr}\left(\nabla\xi(x)^2 + A(x, \xi(x))\nabla\xi(x) + B(x, \xi(x))\right)$$

for vector fields $\xi \in C^2(\mathbb{R}^d; \mathbb{R}^d)$. The maps A and B are defined in Lemma 5.1. We prove that this generalized curvature is independent of the choice of coordinates (Theorem 5.7). In the case where ξ take a special form (that naturally arises from the optimal transport problem), we provide an explicit formula for $\mathcal{K}_x$ in Theorem 5.8. Lastly, we furnish an example of a Lagrangian cost with non-negative generalized curvature that is not given by squared Riemannian distance. This induces a geometry on the L-Wasserstein space where the generalized entropy functional (4.1) is displacement convex along suitable curves.

This paper is organized as follows: In the first four sections, we will review the optimal transport problem induced by a Tonelli Lagrangian, up to and including the notion of displacement convexity. The thesis of Schachter [19] provides a good overview of key definitions and results needed. Section 2 covers some basic notation. Section 3 reviews some ideas from [19]; chief among them is the relationship between the various formulations of the optimal transport problem. Section 4 discusses functionals along curves in Wasserstein space, including a computation of the displacement Hessian. Section 5 introduces the definition and various properties of the generalized curvature $\mathcal{K}_x$. Lastly, Sect. 6 provides an example of a Lagrangian with everywhere non-negative generalized curvature.

2 Notation

We will take our underlying manifold to be $M = \mathbb{R}^d$ and identify its tangent bundle $T\mathbb{R}^d \cong \mathbb{R}^d \times \mathbb{R}^d$. Let $\mathcal{P}^{ac} = \mathcal{P}^{ac}(\mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d$ that are absolutely continuous with respect to the $d-$dimensional Lebesgue measure (denoted $\mathcal{L}^d$). An element of $\mathcal{P}^{ac}$ will often be identified by its density ρ. Given $\rho \in \mathcal{P}^{ac}$ and a measurable function $T : \mathbb{R}^d \to \mathbb{R}^d$, $T_{\#}\rho$ will denote the push-forward measure of ρ.

Definition 2.1 (*Tonelli Lagrangian*) A function $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called a *Tonelli Lagrangian* if it satisfies the following conditions:

(i) $L \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$.

(ii) For every $x \in \mathbb{R}^d$, the function $L(x, \cdot) : \mathbb{R}^d \to \mathbb{R}$ is strictly convex.
(iii) L has asymptotic superlinear growth in the variable v, in the sense that there exists a constant $c_0 \in \mathbb{R}$ and a function $\theta : \mathbb{R}^d \to [0, +\infty)$ with

$$\lim_{|v| \to +\infty} \frac{\theta(v)}{|v|} = +\infty$$

such that

$$L(x, v) \geq c_0 + \theta(v) \tag{2.1}$$

for all $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$.

Throughout this manuscript, $L \in C^k(\mathbb{R}^d \times \mathbb{R}^d)$, $k \geq 3$ will be assumed to be a Tonelli Lagrangian and we will work with the underlying space $(\mathbb{R}^d, L)$. We denote the gradient with respect to the x (position) and v (velocity) variables by $\nabla_x L, \nabla_v L \in \mathbb{R}^d$ respectively. Similarly, the second-order derivatives will be denoted by $\nabla^2_{xx} L$, $\nabla^2_{vv} L$, $\nabla^2_{xv} L$, $\nabla^2_{vx} L = \nabla^2_{xv} L^\top \in \mathbb{R}^{d\times d}$. We will assume that the Hessian $\nabla^2_{vv} L(x, v)$ is positive-definite for every $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$. The time derivative of a function $f(t)$ will be denoted by $\dot{f} = \frac{df}{dt}$.

3 Optimal transport problem induced by a Tonelli Lagrangian

3.1 Lagrangian optimal transport problem

The goal of this section is to establish the different formulations of the optimal transport problem with cost induced by a Tonelli Lagrangian L. In this first subsection, the Lagrangian optimal transport problem will be presented. We will also briefly recall the classical Monge–Kantorovich theory. Most of the material in the subsection can be found in [5, 11, 19, 21]. In subsection 3.2 we will present an Eulerian perspective and its connections to viscosity solutions of the *Hamilton–Jacobi equation*.

Definition 3.1 (*Action functional*) Let $T > 0$ and $\gamma \in W^{1,1}([0, T]; \mathbb{R}^d)$ be a curve. The *action of* L *on* γ is

$$\mathcal{A}_{L,T}(\gamma) = \int_0^T L(\gamma(t), \dot{\gamma}(t))\, dt. \tag{3.1}$$

This induces a *cost function* $c_{L,T} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ given by

$$c_{L,T}(x, y) = \inf\{\mathcal{A}_{L,T}(\gamma) \ : \ \gamma \in W^{1,1}([0, T]; \mathbb{R}^d), \gamma(0) = x, \gamma(T) = y\}. \tag{3.2}$$

A curve γ with $\gamma(0) = x, \gamma(T) = y$ is called an *action-minimizing curve from* x *to* y if $\mathcal{A}_{L,T}(\gamma) = c_{L,T}(x, y)$.

Theorem 3.2 ([11] Appendix B) *For any* $x, y \in \mathbb{R}^d$, *there exists an action-minimizing curve* γ *from* x *to* y *such that*

(i) $\mathcal{A}_{L,T}(\gamma) = c_{L,T}(x, y)$
(ii) $\gamma \in C^k([0, T]; \mathbb{R}^d)$
(ii) γ *satisfies the Euler–Lagrange equation*

$$\frac{d}{dt}((\nabla_v L)(\gamma, \dot{\gamma})) = (\nabla_x L)(\gamma, \dot{\gamma}) \tag{3.3}$$

Definition 3.3 (*Lagrangian flow*) The Lagrangian flow $\Phi : [0, +\infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ is defined by

$$\begin{cases} \frac{d}{dt}((\nabla_v L)(\Phi)) = (\nabla_x L)(\Phi) \\ \Phi(0, x, v) = (x, v) \end{cases}$$

We refer the reader to [11] and [19] for further properties of the cost function $c_{L,T}$. In particular, it is locally Lipschitz and thus differentiable almost everywhere by Rademacher's theorem. Moreover, if either $\frac{\partial}{\partial x} c_{L,T}(x_0, y_0)$ or $\frac{\partial}{\partial y} c_{L,T}(x_0, y_0)$ exists at (x_0, y_0), then the action-minimizing curve from x_0 to y_0 is unique. With the cost $c_{L,T}$, we may state the Monge problem and the Kantorovich problem.

Definition 3.4 (*Monge problem*) Let $\rho_0, \rho_T \in \mathcal{P}^{ac}$. The Monge optimal transport problem from ρ_0 to ρ_T for the cost $c_{L,T}$ is the minimization problem

$$\inf_M \left\{ \int_{\mathbb{R}^d} c_{L,T}(x, M(x))\rho_0(x)\, dx \ : \ M_\# \rho_0 = \rho_T \ , \ M \text{ Borel measurable} \right\}. \tag{3.4}$$

Definition 3.5 (*Kantorovich problem*) Let $\Pi(\rho_0, \rho_T)$ denote the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals ρ_0 and ρ_T. Then the Kantorovich optimal transport problem from ρ_0 to ρ_T for the cost $c_{L,T}$ is the minimization problem

$$\inf_\pi \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c_{L,T}(x, y)\, d\pi(x, y) \ : \ \pi \in \Pi(\rho_0, \rho_T) \right\}. \tag{3.5}$$

A minimizer π is called an optimal transport plan. The infimum in (3.5) is denoted $W_{c_{L,T}}(\rho_0, \rho_T)$ and it is called the *Kantorovich cost from* ρ_0 *to* ρ_T.

If $W_{c_{L,T}}(\rho_0, \rho_T)$ is finite, then the Monge problem with cost $c_{L,T}$ admits an optimizer M (called the Monge map) that is unique ρ_0-almost everywhere [11]. Note that the Monge problem is only concerned with the initial and final states (i.e. ρ_0, ρ_T). To interpolate between ρ_0 and ρ_T in a way that respects the cost $c_{L,T}$, we consider the Lagrangian formulation of the optimal transport problem induced by L.

Definition 3.6 (*Lagrangian optimal transport problem*) Let $\rho_0, \rho_T \in \mathcal{P}^{ac}$. The Lagrangian optimal transport problem from ρ_0 to ρ_T induced by the Tonelli Lagrangian L is the minimization problem

$$\inf_\sigma \left\{ \int_0^T \int_{\mathbb{R}^d} L(\sigma(t, x), \dot{\sigma}(t, x))\rho_0(x)\, dx\, dt \right\} \tag{3.6}$$

where the infimum is taken over all $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that

(i) $\sigma(\cdot, x) \in W^{1,1}([0, T]; \mathbb{R}^d)$ for every $x \in \mathbb{R}^d$
(ii) $\sigma(t, \cdot)$ is Borel measurable for every $t \in [0, T]$
(iii) $\sigma(0, x) = x$ for every $x \in \mathbb{R}^d$
(iv) $\sigma(T, \cdot)_\# \rho_0 = \rho_T$

In [19], it is shown that if $W_{c_{L,T}}(\rho_0, \rho_T)$ is finite, then the Lagrangian optimal transport problem admits an optimizer σ such that $\sigma(\cdot, x)$ is an action-minimizing curve from $\sigma(0, x) = x$ to $\sigma(T, x)$ for every $x \in \mathbb{R}^d$. Moreover, the map $\sigma(T, \cdot)$ coincides with the Monge map M and so is unique ρ_0-almost everywhere. With an optimizer σ, we can define the notion of *displacement interpolation*, which is the analogue of a geodesic in $\mathcal{P}^{ac}$.

Definition 3.7 (Displacement interpolant) Let $\rho_0, \rho_T \in \mathcal{P}^{ac}$ be such that the Kantorovich cost $W_{c_{L,T}}(\rho_0, \rho_T)$ is finite. Let σ be an optimizer of the Lagrangian optimal transport problem. Then the *displacement interpolant* between ρ_0 and ρ_T for the cost $c_{L,T}$ is the measure-valued map

$$[0, T] \ni t \mapsto \mu_t = \sigma(t, \cdot)_\# \rho_0.$$

Since μ_t is absolutely continuous with respect to $\mathcal{L}^d$ for every $t \in [0, T]$ ([11] Theorem 5.1), we will also identify μ_t with its density ρ_t. Subsequently, we will always denote a displacement interpolant by a function $\rho : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ and use the notation $\rho_t = \rho(t, \cdot)$ whenever the intention is clear. Since the maps $\sigma(t, \cdot)$ are uniquely defined (ρ_0-almost everywhere) on the support of ρ_0, the displacement interpolant is well-defined. Moreover, the map $\sigma\big|_{[0,t]\times\mathbb{R}^d}$ for an intermediary time $t \in [0, T]$ optimizes the Lagrangian optimal transport problem from ρ_0 to ρ_t, i.e.

$$W_{c_{L,t}}(\rho_0, \rho_t) = \int_0^t \int_{\mathbb{R}^d} L(\sigma(s, x), \dot{\sigma}(s, x))\rho_0(x)\, dx\, ds.$$

In order to discuss the Eulerian formulation of the optimal transport problem, we need to introduce the *Kantorovich duality*. We do so in accordance with the convention of [21].

Theorem 3.8 (Kantorovich duality) *The Kantorovich optimal transport problem from ρ_0 to ρ_T for the cost $c_{L,T}$ has a dual formulation*

$$\begin{aligned} &\inf_\pi \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} c_{L,T}(x, y)\, d\pi(x, y) \;:\; \pi \in \Pi(\rho_0, \rho_T) \right\} \\ &= \sup_{(u_0,u_T)} \left\{ \int_{\mathbb{R}^d} u_T(y)\rho_T(y)\, dy - \int_{\mathbb{R}^d} u_0(x)\rho_0(x)\, dx \;:\; (u_0, u_T) \in L^1(\rho_0) \times L^1(\rho_T)\,, \right. \\ &\qquad \left. u_T(y) - u_0(x) \le c_{L,T}(x, y) \quad \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \right\} \end{aligned}$$

Moreover, we may assume that

$$\begin{aligned} u_T(y) &= \inf_{x\in\mathbb{R}^d} \left\{u_0(x) + c_{L,T}(x, y)\right\} \\ u_0(x) &= \sup_{y\in\mathbb{R}^d} \left\{u_T(y) - c_{L,T}(x, y)\right\} \end{aligned}$$

If (u_0, u_T) is an optimizer of the dual problem, then u_0 and u_T are called Kantorovich potentials.

Remark 3.9 If the Monge optimal transport problem from ρ_0 to ρ_T for the cost $c_{L,T}$ admits a minimizer M (unique ρ_0-almost everywhere), then any optimal transport plan $\pi \in \Pi(\rho_0, \rho_T)$ is concentrated on the graph of M [11]. Moreover, if u_0 and u_T are Kantorovich potentials, then

$$u_T(y) - u_0(x) \le c_{L,T}(x, y)$$

for every $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ and we have equality

$$u_T(M(x)) - u_0(x) = c_{L,T}(x, M(x))$$

for x ρ_0-almost everywhere (see [21] Theorem 5.10).

3.2 Eulerian formulation

The paper by Benamou and Brenier [4] is one of the earliest works establishing the Eulerian formulation and its connection to Hamilton–Jacobi equations. Subsequently, the relationships between the different formulations of the optimal transport problem were further studied (for instance, [5]).

In particular, the Eulerian view establishes the displacement interpolant as a solution to the *continuity equation*. First, we state some basic facts about the *Hamiltonian*.

The Hamiltonian associated with the Tonelli Lagrangian L is defined as the *Legendre transform* of L with respect to the variable v, i.e.

$$H(x, p) = \sup_{v\in\mathbb{R}^d} \{\langle p, v\rangle - L(x, v)\}. \tag{3.7}$$

Thus, the Hamiltonian H satisfies the *Fenchel-Young inequality*

$$\langle v, p\rangle \le H(x, p) + L(x, v) \tag{3.8}$$

for all $x, v, p \in \mathbb{R}^d$, with equality if and only if

$$p = (\nabla_v L)(x, v). \tag{3.9}$$

Moreover, $H \in C^k(\mathbb{R}^d \times \mathbb{R}^d)$ and

$$(\nabla_v L)(x, (\nabla_p H)(x, r)) = (\nabla_p H)(x, (\nabla_v L)(x, r)) = r. \tag{3.10}$$

Let $u_0 : \mathbb{R}^d \to [-\infty, +\infty]$ be a function and $T > 0$. We define the *Lax-Oleinik evolution* $u : [0, T] \times \mathbb{R}^d \to [-\infty, +\infty]$ of u_0 by

$$\begin{aligned} u(t, x) &:= \inf_{\gamma} \left\{ u_0(\gamma(0)) + \int_0^t L(\gamma(\tau), \dot\gamma(\tau))\, d\tau \ : \ \gamma \in W^{1,1}([0, t]; \mathbb{R}^d)\ ,\ \gamma(t) = x \right\} \\ &= \inf_{\gamma} \left\{ u_0(\gamma(0)) + \mathcal{A}_{L,t}(\gamma) \ : \ \gamma \in W^{1,1}([0, t]; \mathbb{R}^d)\ ,\ \gamma(t) = x \right\} \\ &= \inf_{y\in\mathbb{R}^d} \left\{ u_0(y) + c_{L,t}(y, x) \right\} \end{aligned} \tag{3.11}$$

so that $u(0, x) = u_0(x)$.

Remark 3.10 Since L is bounded below, if there exists some $(t^*, x^*) \in (0, T] \times \mathbb{R}^d$ such that $u(t^*, x^*)$ is finite, then u is finite on all of $[0, T] \times \mathbb{R}^d$.

It is known that if u is finite, then it is a viscosity solution of the Hamilton–Jacobi equation

$$\frac{\partial u}{\partial t} + H(x, \nabla u) = 0 \tag{3.12}$$

(see [9] Section 7.2 and [10] Theorem 1.1).

Definition 3.11 (*Calibrated curve*) Let $f : [t_0, t_1] \times \mathbb{R}^d$ be a function. A curve $\gamma \in W^{1,1}([t_0, t_1]; \mathbb{R}^d)$ is called a *$(f, L)-$calibrated curve* if $f(t_0, \gamma(t_0))$, $f(t_1, \gamma(t_1))$ and $\int_{t_0}^{t_1} L(\gamma(t), \dot\gamma(t))\, dt$ are all finite and

$$f(t_1, \gamma(t_1)) - f(t_0, \gamma(t_0)) = \int_{t_0}^{t_1} L(\gamma(t), \dot\gamma(t))\, dt. \tag{3.13}$$

In the following proposition, we mention some properties of u that are of interest to us. The proofs can be found in [6, 9, 10].

Proposition 3.12 *Let u be defined as in (3.11). If u is finite, then the following hold:*

(i) *u is continuous and locally semi-concave on $(0, T) \times \mathbb{R}^d$.*
(ii) *u is a viscosity solution of the Hamilton–Jacobi equation*

$$\frac{\partial u}{\partial t} + H(x, \nabla u) = 0.$$

(iii) *If $[a, b] \subset [0, T]$ and $\gamma : [a, b] \to \mathbb{R}^d$ is a $(u, L)-$calibrated curve, then u is differentiable at $(t, \gamma(t))$ for every $t \in [a, b]$ and we have*

$$\nabla u(t, \gamma(t)) = (\nabla_v L)(\gamma(t), \dot{\gamma}(t)). \tag{3.14}$$

(iv) *If u is differentiable at (t^*, x^*), then there is at most one $(u, L)-$calibrated curve $\gamma : [a, b] \to \mathbb{R}^d$ with $t^* \in [a, b]$ and $\gamma(t^*) = x^*$.*

We now return to the optimal transport problem from $\rho_0 \in \mathcal{P}^{ac}$ to $\rho_T \in \mathcal{P}^{ac}$ induced by L. Suppose that $W_{c_{L,T}}(\rho_0, \rho_T)$ is finite and let $u_0 \in L^1(\rho_0)$ be a Kantorovich potential.

Proposition 3.13 *Let $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be an optimizer of the Lagrangian optimal transport problem from ρ_0 to ρ_T induced by L. Let (u_0, u_T) be the corresponding Kantorovich potentials and $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ be the Lax–Oleinik evolution of u_0. Then $(\nabla u)(t, \sigma(t, x))$ exists for all $t \in [0, T]$ and x ρ_0-almost everywhere. In addition, σ satisfies the relation*

$$\dot{\sigma}(t, x) = (\nabla_p H)(\sigma(t, x), (\nabla u)(t, \sigma(t, x))). \tag{3.15}$$

Proof By Remark 3.10, u is finite since $u(T, \cdot) = u_T \in L^1(\rho_T)$. By Remark 3.9, the Kantorovich potentials (u_0, u_T) satisfy

$$\begin{aligned} & u_T(\sigma(T, x)) - u_0(x) = c_{L,T}(x, \sigma(T, x)) \\ \iff & u(T, \sigma(T, x)) - u(0, \sigma(0, x)) = c_{L,T}(x, \sigma(T, x)) \end{aligned}$$

for x ρ_0-almost everywhere (recall that $\sigma(T, \cdot)$ coincides with the Monge map). Thus, for ρ_0-almost every x, the curve $t \mapsto \sigma(t, x)$ is a $(u, L)-$calibrated curve and so $(\nabla u)(t, \sigma(t, x)) = (\nabla_v L)(\sigma(t, x), \dot{\sigma}(t, x))$ exists by Proposition 3.12. Using identity (3.10), we get

$$\dot{\sigma}(t, x) = (\nabla_p H)(\sigma(t, x), (\nabla u)(t, \sigma(t, x))).$$

□

Remark 3.14 Let $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be a time-dependent vector field that agrees with $(\nabla_p H)(x, (\nabla u)(t, x))$ on the set

$$S_t := \{\sigma(t, y) \in \mathbb{R}^d \ : \ y \in \text{supp}(\rho_0)\ ,\ (\nabla u)(t, \sigma(t, y)) \text{ exists}\}$$

for each $t \in [0, T]$. Using the definition of the displacement interpolant $\rho_t = \sigma(t, \cdot)_\# \rho_0$, and the fact that $(\nabla u)(t, \sigma(t, x))$ exists for all $t \in [0, T]$ and ρ_0-almost every $x \in \mathbb{R}^d$, we have that the set

$$\{\sigma(t, y) \in \mathbb{R}^d \ : \ y \in \text{supp}(\rho_0)\ ,\ u \text{ not differentiable at } (t, \sigma(t, y))\}$$

is a set of zero ρ_t-measure. Thus, S_t has full ρ_t-measure and so $V(t,x) = (\nabla_p H)(x, (\nabla u)(t,x))$ ρ_t-almost everywhere. By (3.15), $\dot{\sigma}(t,x) = V(t, \sigma(t,x))$ for all $t \in [0,T]$ and ρ_0-almost every $x \in \mathbb{R}^d$. This means that ρ_t and V solve the continuity equation

$$\frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t V) = 0 \tag{3.16}$$

in the sense of distributions ([19] Proposition 3.4.3).

4 Generalized entropy functional and displacement Hessian

Otto calculus and Schachter's Eulerian calculus both allow for explicit computations, assuming that all relevant quantities possess sufficient regularity. However, the regularity of a displacement interpolant ρ depends on the Lagrangian L, the initial and final densities (ρ_0, ρ_T), and the optimal trajectories σ (or the velocity field V in the Eulerian framework). In general, the Kantorovich potential u_0 arising from an optimal transport problem induced by a Tonelli Lagrangian L is only known to be semiconcave, differentiable $\mathcal{L}^d-$almost everywhere, and its gradient ∇u_0 is only locally bounded (see [13] and [14] Appendix C). This implies that the initial velocity $V(0,x) = (\nabla_p H)(x, \nabla u_0(x))$ is only locally bounded. As such, even if the initial density ρ_0 is smooth, its regularity may fail to propagate along the displacement interpolant.

For our purpose of computing displacement Hessians, we require displacement interpolants to be of class C^2. Fortunately, such displacement interpolants do exist and we can construct them if we impose two additional criteria on L.

4.1 C^2 displacement interpolants

Let $L \in C^{k+1}(\mathbb{R}^d \times \mathbb{R}^d)$, $k \geq 3$ be a Tonelli Lagrangian satisfying two additional criteria (see [6] Chapters 6.3, 6.4).

(L1) There exists $\tilde{c}_0 \geq 0$ and $\tilde{\theta} : [0,\infty) \to [0,\infty)$ with

$$\lim_{r \to +\infty} \frac{\tilde{\theta}(r)}{r} = +\infty$$

such that

$$L(x,v) \geq \tilde{\theta}(|v|) - \tilde{c}_0.$$

In addition, $\tilde{\theta}$ is such that for any $M > 0$ there exists $K_M > 0$ with

$$\tilde{\theta}(r+m) \leq K_M[1 + \tilde{\theta}(r)]$$

for all $m \in [0, M]$ and all $r \geq 0$.

(L2) For any $r > 0$, there exists $C_r > 0$ such that

$$|(\nabla_x L)(x,v)| + |(\nabla_v L)(x,v)| < C_r \tilde{\theta}(|v|)$$

for all $|x| \leq r$, $v \in \mathbb{R}^d$.

Some common examples of Tonelli Lagrangians satisfying these criteria include the *Riemannian kinetic energy*

$$L(x,v) = \frac{1}{2} g_x(v,v)$$

where g_x denotes the underlying Riemannian metric tensor, and Lagrangians that arise from mechanics

$$L(x, v) = \frac{1}{2} g_x(v, v) + U(x)$$

for some appropriate potential $U : \mathbb{R}^d \to \mathbb{R}$.

Let H be the corresponding Hamiltonian.

Lemma 4.1 *Let $u_0 \in C^{k+1}(\mathbb{R}^d)$ with $u_0(x) \geq -\tilde{c}_0$ for all $x \in \mathbb{R}^d$. Let $u : [0, +\infty) \times \mathbb{R}^d \to [-\infty, +\infty]$ be the Lax–Oleinik evolution of u_0, as defined in (3.11). For $x \in \mathbb{R}^d$, consider the Lagrangian flow (introduced in Definition 3.3)*

$$\Phi(t, x, V(0, x)) = (\Phi_1(t, x, V(0, x)), \Phi_2(t, x, V(0, x))) \ , \quad t \in [0, +\infty)$$

where $V : [0, +\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ is a time-dependent vector field defined by

$$V(t, x) = (\nabla_p H)(x, (\nabla u)(t, x)).$$

(Here, Φ_1 and Φ_2 are the x and v components of Φ respectively.) If we let $\sigma(t, x) = \Phi_1(t, x, V(0, x))$, then $\dot{\sigma}(t, x) = V(t, \sigma(t, x))$ for all $t \in [0, +\infty)$, $x \in \mathbb{R}^d$. Moreover, $\sigma(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is a C^k-diffeomorphism for every $t \in [0, +\infty)$.

Proof Since L and u_0 are both bounded below, we have

$$\begin{aligned} u(t, x) &= \inf_\gamma \left\{ u_0(\gamma(0)) + \int_0^t L(\gamma(\tau), \dot{\gamma}(\tau))\, d\tau \ , \ \gamma(t) = x \right\} \\ &\geq -\tilde{c}_0 - \tilde{c}_0 t \\ &> -\infty \end{aligned}$$

and so u is finite. From [10], u is a continuous viscosity solution of the Hamilton–Jacobi equation (3.12) and we know that for each $(t, x) \in (0, +\infty) \times \mathbb{R}^d$, there exists a unique $(u, L)-$calibrated curve $\gamma_x : [0, t] \to \mathbb{R}^d$ such that $\gamma_x(t) = x$. Moreover, $(\nabla u)(s, \gamma_x(s))$ exists for all $s \in [0, t]$ and is given by

$$\begin{aligned} & (\nabla u)(s, \gamma_x(s)) = (\nabla_v L)(\gamma_x(s), \dot{\gamma}_x(s)) \\ \iff & \dot{\gamma}_x(s) = (\nabla_p H)(\gamma_x(s), (\nabla u)(s, \gamma_x(s))) \\ \iff & \dot{\gamma}_x(s) = V(s, \gamma_x(s)) \end{aligned}$$

Since each γ_x is necessarily an action-minimizing curve from $\gamma_x(0)$ to $\gamma_x(t) = x$, it is the unique solution curve to the Euler–Lagrange system

$$\begin{cases} \frac{d}{dt}((\nabla_v L)(\gamma, \dot{\gamma})) = (\nabla_x L)(\gamma, \dot{\gamma}) \\ \gamma(0) = \gamma_x(0) \\ \dot{\gamma}(0) = \dot{\gamma}_x(0) \end{cases}$$

Therefore, $\sigma(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is a bijection for all $t \in [0, +\infty)$ and $\dot{\sigma}(t, x) = V(t, \sigma(t, x))$ for all $(t, x) \in [0, +\infty) \times \mathbb{R}^d$.

Lastly, since $L \in C^{k+1}(\mathbb{R}^d \times \mathbb{R}^d)$ and $u_0 \in C^{k+1}(\mathbb{R}^d)$, we have that $u \in C^{k+1}([0, +\infty) \times \mathbb{R}^d)$ [6]. As $\nabla_p H \in C^k(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$, we have $V(t, \cdot) \in C^k(\mathbb{R}^d; \mathbb{R}^d)$ and so $\sigma(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is a C^k-diffeomorphism for every $t \in [0, +\infty)$. □

Proposition 4.2 *Let $\rho_0 \in \mathcal{P}^{ac} \cap C_c^2(\mathbb{R}^d)$ be a compactly supported density. Then for any $T > 0$, there exists a C^2 displacement interpolant $\rho : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ with $\rho(0, \cdot) = \rho_0$.*

Proof Let u_0, u, V, σ be defined as in Lemma 4.1 and fix $T > 0$. For $t \in [0, T]$, define

$$\rho(t, \cdot) = \sigma(t, \cdot)_{\#}\rho_0.$$

We claim that σ is an optimizer of the Lagrangian optimal transport problem from ρ_0 to $\rho_T = \rho(T, \cdot)$, which would imply that ρ is indeed a displacement interpolant. Let $\phi : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy the four conditions in Definition 3.6. By Lemma 4.1, $t \mapsto \sigma(t, x)$ is a $(u, L)-$calibrated curve for each $x \in \mathbb{R}^d$. Thus, for every $x \in \mathbb{R}^d$,

$$\begin{aligned} u(T, \sigma(T, x)) - u(0, \sigma(0, x)) &= \int_0^T L(\sigma(t, x), \dot{\sigma}(t, x))\, dt \\ \iff u(T, \sigma(T, x)) - u_0(x) &= \int_0^T L(\sigma(t, x), \dot{\sigma}(t, x))\, dt \\ \implies \int_{\mathbb{R}^d} [u(T, \sigma(T, x)) - u_0(x)]\rho_0(x)\, dx &= \int_{\mathbb{R}^d} \int_0^T L(\sigma(t, x), \dot{\sigma}(t, x))\rho_0(x)\, dt\, dx \end{aligned}$$

By the definition of pushforward measure, the LHS of the last equality is

$$\begin{aligned} \int_{\mathbb{R}^d} [u(T, \sigma(T, x)) - u_0(x)]\rho_0(x)\, dx &= \int_{\mathbb{R}^d} u(T, y)\rho_T(y)\, dy - \int_{\mathbb{R}^d} u_0(x)\rho_0(x)\, dx \\ &= \int_{\mathbb{R}^d} [u(T, \phi(T, x)) - u(0, \phi(0, x))]\rho_0(x)\, dx \end{aligned}$$

where the last equality is due to the assumption that $\phi(T, \cdot)_{\#}\rho_0 = \rho_T$ and $\phi(0, x) = x$. By the definition of u (i.e. (3.11)), we have that

$$\begin{aligned} u(T, \phi(T, x)) - u(0, \phi(0, x)) &\le \int_0^T L(\phi(t, x), \dot{\phi}(t, x))\, dt \\ \implies \int_{\mathbb{R}^d} [u(T, \phi(T, x)) - u(0, \phi(0, x))]\rho_0(x)\, dx &\le \int_{\mathbb{R}^d} \int_0^T L(\phi(t, x), \dot{\phi}(t, x))\rho_0(x)\, dt\, dx \end{aligned}$$

for every $x \in \mathbb{R}^d$. Thus,

$$\int_{\mathbb{R}^d} \int_0^T L(\sigma(t, x), \dot{\sigma}(t, x))\rho_0(x)\, dt\, dx \le \int_{\mathbb{R}^d} \int_0^T L(\phi(t, x), \dot{\phi}(t, x))\rho_0(x)\, dt\, dx.$$

Since ϕ was arbitrary, σ is indeed an optimizer of the Lagrangian optimal transport problem from ρ_0 to ρ_T.

By Lemma 4.1, $\sigma(t, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is a C^k-diffeomorphism for every $t \in [0, T]$ and $\sigma(\cdot, x) \in C^{k+1}([0, T]; \mathbb{R}^d)$ for every $x \in \mathbb{R}^d$. Using the change-of-variables formula,

$$\begin{aligned} \rho(t, y) &= \frac{\rho_0(\cdot)}{|\det \nabla \sigma(t, \cdot)|}\bigg|_{[\sigma(t,\cdot)]^{-1}(y)} \\ &= \frac{\rho_0(\cdot)}{\det \nabla \sigma(t, \cdot)}\bigg|_{[\sigma(t,\cdot)]^{-1}(y)} \end{aligned}$$

where $\det \nabla \sigma(t, \cdot) > 0$ because $\sigma(0, x) = x \implies \det \nabla \sigma(0, \cdot) = 1$. Since $k \ge 3$, $\rho \in C^2([0, T] \times \mathbb{R}^d)$. □

4.2 Displacement Hessian

Let $F \in C^2((0, +\infty)) \cap C([0, +\infty))$ be a function satisfying

(F1) $F(0) = 0$,
(F2) $s^2 F''(s) \geq sF'(s) - F(s) \geq 0, \quad \forall s \in [0, +\infty)$.

If $\rho_0 \in \mathcal{P}^{ac}$ is such that $F(\rho_0) \in L^1(\mathbb{R}^d)$, we define the *generalized entropy functional*

$$\mathcal{F}(\rho_0) = \int_{\mathbb{R}^d} F(\rho_0(x))\, dx. \tag{4.1}$$

This is well-defined at least on $\mathcal{P}^{ac} \cap C_c^0(\mathbb{R}^d)$ since $F(0) = 0$ implies

$$\int_{\mathbb{R}^d} F(\rho_0(x))\, dx = \int_{\text{supp}(\rho_0)} F(\rho_0(x))\, dx$$

which is finite.

Remark 4.3 If ρ_0 is the density of a fluid and $\mathcal{F}(\rho_0)$ is the internal energy, then $\rho_0 F'(\rho_0) - F(\rho_0)$ can be interpreted as a pressure [12, 21].

Definition 4.4 (*Displacement convexity*) The generalized entropy functional $\mathcal{F}$ is said to be convex along a displacement interpolant ρ_t, $t \in [0, T]$, if $\mathcal{F}(\rho_t)$ is finite and

$$\mathcal{F}(\rho_t) \leq \frac{T-t}{T}\mathcal{F}(\rho_0) + \frac{t}{T}\mathcal{F}(\rho_T) \tag{4.2}$$

for every $t \in [0, T]$. $\mathcal{F}$ is said to be *displacement convex* if it is convex along every displacement interpolant (on which $\mathcal{F}$ is real-valued).

Remark 4.5 When the displacement interpolant is a "straight line", McCann proved that $\mathcal{F}$ is displacement convex if $s \mapsto s^d F(s^{-d})$ is convex and non-increasing on $(0, +\infty)$ [16]. In this context, a "straight line" displacement interpolant refers to one of the form

$$\rho_t = \left(\frac{T-t}{T}\text{id} + \frac{t}{T}M\right)_{\#}\rho_0.$$

where M is the Monge map between ρ_0 and ρ_T.

Along a suitable displacement interpolant ρ_t, if the map $t \mapsto \mathcal{F}(\rho_t)$ is C^2, then the condition that $\frac{d^2}{dt^2}\mathcal{F}(\rho_t) \geq 0$ ensures convexity of $\mathcal{F}$ along ρ_t. The following displacement Hessian formula is a special case of Theorem 4.3.2 of [19].

Theorem 4.6 (Displacement Hessian formula) *Let $\rho \in C^2([0, T] \times \mathbb{R}^d)$ be a displacement interpolant, with $\rho_0 = \rho(0, \cdot)$ compactly supported. Let $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be an optimizer of the Lagrangian optimal transport problem from ρ_0 to ρ_T. Let $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be defined as in Remark 3.14 so that ρ, V satisfy the continuity equation $\dot{\rho} = -\nabla \cdot (\rho V)$. Assume that σ and V are C^2 at least on the set*

$$\bigcup_{t \in [0,T]} \{t\} \times \text{supp}(\rho_t).$$

Then $\frac{d^2}{dt^2}\mathcal{F}(\rho)$ exists for every $t \in [0, T]$ and is given by

$$\frac{d^2}{dt^2}\mathcal{F}(\rho) = \int_{\mathbb{R}^d} (\rho G'(\rho) - G(\rho))(\nabla \cdot V)^2 + G(\rho)(\text{tr}((\nabla V)^2) - \nabla \cdot W)\, dx \tag{4.3}$$

where $G : [0, +\infty) \to \mathbb{R}$ *is defined by*

$$G(s) = sF'(s) - F(s) \tag{4.4}$$

$$G'(s) = sF''(s) \tag{4.5}$$

and

$$W = \dot{V} + \nabla V V. \tag{4.6}$$

Remark 4.7 The requirement that ρ_0 is compactly supported serves to ensure that $\mathcal{F}$ is finite along ρ. In addition, the compactness of $\text{supp}(\rho_0)$ and the continuity of σ together ensures that the set $\{\sigma(t, x) \ : \ t \in [0, T]\ ,\ x \in \text{supp}(\rho_0)\}$ is compact. Thus,

$$\Sigma := \bigcup_{t\in[0,T]} \text{supp}(\rho_t)$$

is bounded, up to a set of zero $\mathcal{L}^d$−measure. This means that $\frac{d^2}{dt^2}\mathcal{F}(\rho)$ exists for every $t \in [0, T]$ and satisfies

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{F}(\rho) &= \frac{d^2}{dt^2}\int_\Sigma F(\rho(t, x))\, dx\\ &= \int_\Sigma \frac{d^2}{dt^2} F(\rho(t, x))\, dx.\end{aligned}$$

Remark 4.8 By Remark 3.14, for every $t \in [0, T]$, $V(t, \cdot)$ is uniquely defined on $\text{supp}(\rho_t)$ ρ_t−almost everywhere. Thus, (4.3) is well-defined.

Proof The displacement Hessian is

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{F}(\rho) &= \int F''(\rho)\dot{\rho}^2 + F'(\rho)\ddot{\rho}\, dx\\ &= \int F''(\rho)\dot{\rho}^2 - F'(\rho)\nabla \cdot (\dot{\rho}V + \rho\dot{V})\, dx\end{aligned}$$

Integrating by parts, the above expression becomes

$$\begin{aligned}&\int F''(\rho)\dot{\rho}^2 + \langle\nabla(F'(\rho)), \dot{\rho}V + \rho\dot{V}\rangle\, dx\\ &= \int F''(\rho)\Big(\dot{\rho}^2 + \langle\nabla\rho, \dot{\rho}V + \rho\dot{V}\rangle\Big)\, dx.\end{aligned}$$

Using the continuity equation $\dot{\rho} = -\nabla \cdot (\rho V)$, the definitions of W and G, and integration by parts, this integral can be written as

$$\int \rho G'(\rho)(\nabla \cdot V)^2 - G(\rho)\nabla \cdot \Big((\nabla \cdot V)V - \nabla V V + W\Big)\, dx.$$

A straightforward computation then yields the desired formula. □

Remark 4.9 Recall that $\rho G'(\rho) - G(\rho) = \rho^2 F''(\rho) - \rho F'(\rho) + F'(\rho) \geq 0$ and $G(\rho) = \rho F'(\rho) - F(\rho) \geq 0$ by assumption (F2). Thus, the condition that $\text{tr}((\nabla V)^2) - \nabla \cdot W \geq 0$ would ensure that $\frac{d^2}{dt^2}\mathcal{F}(\rho) \geq 0$. In the case where the cost is given by squared Riemannian distance, the term $\text{tr}((\nabla V)^2) - \nabla \cdot W$ is a quadratic form involving the Bakry–Emery tensor [19], [21]. In the following section, we will generalize this quadratic form for an arbitrary Tonelli Lagrangian.

5 Generalized curvature for Tonelli Lagrangians

The goal of this section is to define a generalized curvature for the space $(\mathbb{R}^d, L)$. In principle, this generalized curvature is similar to the Ricci curvature in the sense that it is related to the deformation of a shape flowing along action-minimizing curves. The generalized curvature, however, will not be a tensor because it will depend on both the tangent vector and its gradient. Throughout this section, we will assume that L is a C^3 Tonelli Lagrangian.

Let $T > 0$ and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be such that

$$\begin{cases} \frac{d}{dt}\big((\nabla_v L)(\sigma(t,x), \dot\sigma(t,x))\big) = (\nabla_x L)(\sigma(t,x), \dot\sigma(t,x)) \ , \quad \forall (t,x) \in [0,T] \times \mathbb{R}^d \\ \sigma(0,x) = x \ , \quad \forall x \in \mathbb{R}^d \\ \sigma(t,\cdot) : \mathbb{R}^d \to \mathbb{R}^d \text{ is a } C^3 - \text{diffeomorphism for every } t \in [0,T] \end{cases}$$

Let $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ be a time-dependent vector field defined by $\dot\sigma(t, x) = V(t, \sigma(t, x))$ so that $V(t, \cdot) \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ for every $t \in [0, T]$. Following the method outlined in [21] Chapter 14, we first derive Lemma 5.1, which is an ODE of the Jacobian matrix $\nabla\sigma$.

Lemma 5.1 *Define $A, B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ by*

$$A(x,v) = (\nabla^2_{vv}L)(x,v)^{-1}\left[\frac{d}{dt}\big((\nabla^2_{vv}L)(\gamma_{x,v}(t), \dot\gamma_{x,v}(t))\big)\Big|_{t=0} + (\nabla^2_{vx}L)(x,v) - (\nabla^2_{xv}L)(x,v)\right] \tag{5.1}$$

$$B(x,v) = (\nabla^2_{vv}L)(x,v)^{-1}\left[\frac{d}{dt}\big((\nabla^2_{vx}L)(\gamma_{x,v}(t), \dot\gamma_{x,v}(t))\big)\Big|_{t=0} - (\nabla^2_{xx}L)(x,v)\right] \tag{5.2}$$

where $\gamma_{x,v} : [0, \epsilon) \to \mathbb{R}^d$ is the unique curve satisfying the Euler–Lagrange equation with initial conditions $\gamma_{x,v}(0) = x$, $\dot\gamma_{x,v}(0) = v$. Then the Jacobian $\nabla\sigma$ satisfies a second-order matrix equation

$$\nabla\ddot\sigma + A(\sigma,\dot\sigma)\nabla\dot\sigma + B(\sigma,\dot\sigma)\nabla\sigma = 0. \tag{5.3}$$

Proof Taking the spatial gradient of the Euler–Lagrange equation,

$$\begin{aligned} 0 &= \nabla_x\left(\frac{d}{dt}\big((\nabla_v L)(\sigma,\dot\sigma)\big) - (\nabla_x L)(\sigma,\dot\sigma)\right) \\ &= \frac{d}{dt}\left((\nabla^2_{vx}L)(\sigma,\dot\sigma)\nabla\sigma + (\nabla^2_{vv}L)(\sigma,\dot\sigma)\nabla\dot\sigma\right) - (\nabla^2_{xx}L)(\sigma,\dot\sigma)\nabla\sigma - (\nabla^2_{xv}L)(\sigma,\dot\sigma)\nabla\dot\sigma \\ &= \frac{d}{dt}\big((\nabla^2_{vx}L)(\sigma,\dot\sigma)\big)\nabla\sigma + (\nabla^2_{vx}L)(\sigma,\dot\sigma)\nabla\dot\sigma + \frac{d}{dt}\big((\nabla^2_{vv}L)(\sigma,\dot\sigma)\big)\nabla\dot\sigma + (\nabla^2_{vv}L)(\sigma,\dot\sigma)\nabla\ddot\sigma \\ &\quad - (\nabla^2_{xx}L)(\sigma,\dot\sigma)\nabla\sigma - (\nabla^2_{xv}L)(\sigma,\dot\sigma)\nabla\dot\sigma \end{aligned}$$

To conclude, we group by the terms $\nabla\sigma$, $\nabla\dot\sigma$, $\nabla\ddot\sigma$ and multiply by $(\nabla^2_{vv}L)(\sigma,\dot\sigma)^{-1}$. □

Lemma 5.2 *Define*

$$\mathcal{U}(t,x) = (\nabla V)(t, \sigma(t,x)). \tag{5.4}$$

Then

$$\dot{\mathcal{U}} + \mathcal{U}^2 + A(\sigma,\dot\sigma)\mathcal{U} + B(\sigma,\dot\sigma) = 0. \tag{5.5}$$

Proof First, we note that since $\dot{\sigma}(t,x) = V(t,\sigma(t,x))$, we get

$$
\begin{aligned}
&(\nabla\dot{\sigma})(t,x) = (\nabla V)(t,\sigma(t,x))(\nabla\sigma)(t,x) \\
\implies &(\nabla V)(t,\sigma(t,x)) = (\nabla\dot{\sigma})(t,x)((\nabla\sigma)(t,x))^{-1}
\end{aligned}
$$

and so

$$
\mathcal{U}(t,x) = (\nabla\dot{\sigma})(t,x)((\nabla\sigma)(t,x))^{-1}.
$$

Using the matrix identity $\frac{d}{dt}M^{-1} = -M^{-1}\dot{M}M^{-1}$,

$$
\begin{aligned}
\dot{\mathcal{U}} &= (\nabla\ddot{\sigma})(\nabla\sigma)^{-1} - (\nabla\dot{\sigma})(\nabla\sigma)^{-1}(\nabla\dot{\sigma})(\nabla\sigma)^{-1} \\
&= (\nabla\ddot{\sigma})(\nabla\sigma)^{-1} - \mathcal{U}^2.
\end{aligned}
$$

By Lemma 5.1, $\nabla\ddot{\sigma} = -A(\sigma,\dot{\sigma})(\nabla\dot{\sigma}) - B(\sigma,\dot{\sigma})(\nabla\sigma)$ and so

$$
\begin{aligned}
0 &= \dot{\mathcal{U}} + \mathcal{U}^2 + \Big(A(\sigma,\dot{\sigma})(\nabla\dot{\sigma}) - B(\sigma,\dot{\sigma})(\nabla\sigma)\Big)(\nabla\sigma)^{-1} \\
&= \dot{\mathcal{U}} + \mathcal{U}^2 + A(\sigma,\dot{\sigma})\mathcal{U} + B(\sigma,\dot{\sigma}).
\end{aligned}
$$

□

We want to show that the term $\mathrm{tr}((\nabla V)^2) - \nabla\cdot W$ appearing in the displacement Hessian formula (4.3) arises from Eq. (5.5). Taking the trace of (5.5), we have

$$
\frac{d}{dt}\Big((\nabla\cdot V)(t,\sigma)\Big) + \mathrm{tr}\Big((\nabla V)(t,\sigma)^2 + A(\sigma,\dot{\sigma})(\nabla V)(t,\sigma) + B(\sigma,\dot{\sigma})\Big) = 0. \tag{5.6}
$$

On the other hand, direct computation yields

$$
\frac{d}{dt}\Big((\nabla\cdot V)(t,\sigma)\Big) = (\nabla\cdot\dot{V})(t,\sigma) + \langle V(t,\sigma), (\nabla(\nabla\cdot V))(t,\sigma)\rangle.
$$

Since $V(t,\sigma(t,x)) = \dot{\sigma}(t,x)$ and $\sigma(0,x) = x$, we may restate the above equation as

$$
\begin{aligned}
&(\nabla\cdot\dot{V})(t,x) + \langle V(t,x), (\nabla(\nabla\cdot V))(t,x)\rangle \\
&\quad + \mathrm{tr}\Big((\nabla V)(t,x)^2 + A(x,V(t,x))(\nabla V)(t,x) + B(x,V(t,x))\Big) = 0.
\end{aligned} \tag{5.7}
$$

Using the identities

$$
\nabla\cdot((\nabla\cdot V)V) = (\nabla\cdot V)^2 + \langle V, \nabla(\nabla\cdot V)\rangle
$$

and

$$
\dot{V} = -\nabla V V + W,
$$

we see that

$$
\nabla\cdot\dot{V} + \langle V, \nabla(\nabla\cdot V)\rangle = \nabla\cdot(-\nabla V V + W) + \nabla\cdot((\nabla\cdot V)V) - (\nabla\cdot V)^2.
$$

By the computation of the displacement Hessian from the previous section, this is precisely $\mathrm{tr}((\nabla V)^2) - \nabla\cdot W$.

At this point, (5.7) holds for all time-dependent C^2 vector fields whose integral curves satisfy the Euler–Lagrange equation (3.3). To show that (5.7) holds for an arbitrary fixed vector field, we first need to make sense of the term $\dot{V}$ by introducing Definition 5.4.

Proposition 5.3 *Given any fixed vector field $V_0 \in C^2(\mathbb{R}^d;\mathbb{R}^d)$, we may extend it for a short time to a unique time-dependent vector field $V(t,x)$, $t \in [0,\epsilon)$ with the following properties:*

(i) $V(0,\cdot) = V_0$
(ii) *The integral curves of V satisfy the Euler–Lagrange equation, i.e.*

$$\begin{aligned} \dot\sigma(t,x) &= V(t,\sigma(t,x)) \\ \sigma(0,x) &= x \\ \frac{d}{dt}((\nabla_v L)(\sigma,\dot\sigma)) &= (\nabla_x L)(\sigma,\dot\sigma) \end{aligned}$$

Proof We recall Definition 3.3 and the existence of a Lagrangian flow $\Phi = (\Phi_1, \Phi_2)$ satisfying $\frac{d}{dt}((\nabla_v L)(\Phi)) = (\nabla_x L)(\Phi)$. Set $\sigma(t,x) = \Phi_1(t,x,V_0(x))$. The maps $\sigma(t,\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ are defined for all $t \in [0,+\infty)$ and there exists $\epsilon > 0$ such that $\sigma(t,\cdot)$ is invertible for $t \in [0,\epsilon)$. Thus, for $t \in [0,\epsilon)$, we may define the desired vector field by

$$V(t,y) = \dot\sigma(t,\sigma^{-1}(t,y)). \tag{5.8}$$

□

Definition 5.4 Given a Tonelli Lagrangian L, we define the operation

$$\begin{aligned} \Gamma_L : C^2(\mathbb{R}^d;\mathbb{R}^d) &\to C^2(\mathbb{R}^d;\mathbb{R}^d) \\ V_0 &\mapsto \dot V(0,\cdot) \end{aligned}$$

as in Proposition 5.3.

Remark 5.5 By the Euler–Lagrange equation (3.3), we can give an explicit formula for $\Gamma_L(V_0)$. Suppose $\sigma(t,x)$ and $V(t,x)$ satisfy the two properties in Proposition 5.3, then

$$\ddot\sigma(t,x) = \dot V(t,\sigma(t,x)) + (\nabla V)(t,\sigma(t,x))V(t,\sigma(t,x)).$$

Since

$$\ddot\sigma(t,x) = (\nabla^2_{vv}L)(x,V(t,x))^{-1}\Big((\nabla_x L)(x,V(t,x)) - (\nabla^2_{vx}L)(x,V(t,x))V(t,x)\Big)$$

by the Euler–Lagrange equation, we have

$$\begin{aligned} (\Gamma_L(V_0))(x) &= \dot V(0,x) \\ &= (\nabla^2_{vv}L)(x,V_0(x))^{-1}\Big((\nabla_x L)(x,V_0(x)) - (\nabla^2_{vx}L)(x,V_0(x))V_0(x)\Big) \\ &\quad - (\nabla V_0(x))V_0(x). \end{aligned}$$

Definition 5.6 (*Generalized curvature*) Let $\xi \in C^2(\mathbb{R}^d;\mathbb{R}^d)$. For each $x \in \mathbb{R}^d$, we define the *generalized curvature* $\mathcal{K}_x$ by

$$\mathcal{K}_x(\xi) := \mathrm{tr}\Big(\nabla\xi(x)^2 + A(x,\xi(x))\nabla\xi(x) + B(x,\xi(x))\Big) \tag{5.9}$$

where $A, B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are defined as in Lemma 5.1.

Theorem 5.7 *Let $\xi \in C^2(\mathbb{R}^d; \mathbb{R}^d)$. Then*

$$-\big(\nabla \cdot \big(\Gamma_L(\xi)\big)\big)(x) - \langle \xi(x), (\nabla(\nabla \cdot \xi))(x)\rangle = \mathcal{K}_x(\xi). \tag{5.10}$$

In particular, the generalized curvature $\mathcal{K}_x$ is intrinsic, i.e. does not depend on the choice of coordinates.

Proof By Proposition 5.3, we may extend ξ for a short time to a time-dependent vector field $V(t, x)$, with $V(0, \cdot) = \xi$, whose integral curves satisfy the Euler–Lagrange equation. Thus, (5.7) holds for V and we have

$$\begin{aligned}\mathcal{K}_x(\xi) &= \mathcal{K}_x(V(0,\cdot))\\ &\overset{(5.7)}{=} -(\nabla \cdot \dot{V})(0, x) - \langle V(0, x), (\nabla(\nabla \cdot V))(0, x)\rangle\\ &= -\big(\nabla \cdot \big(\Gamma_L(\xi)\big)\big)(x) - \langle \xi(x), (\nabla(\nabla \cdot \xi))(x)\rangle\end{aligned}$$

To show that $\mathcal{K}_x$ is intrinsic, we will show that the operator

$$\xi \mapsto -\nabla \cdot \big(\Gamma_L(\xi)\big) - \langle \xi, \nabla(\nabla \cdot \xi)\rangle \tag{5.11}$$

is invariant under a change of coordinates. By Definition 5.4 and the definition of divergence, $-\nabla \cdot \big(\Gamma_L(\xi)\big)$ is coordinate-free. Next, observe that $\langle \xi, \nabla(\nabla \cdot \xi)\rangle$ is the directional derivative of $\nabla \cdot \xi$ (which is coordinate-free) with respect to ξ. Thus,

$$\langle \xi(x), \nabla(\nabla \cdot \xi)(x)\rangle = \lim_{h \to 0} \frac{(\nabla \cdot \xi)(x + h\xi(x)) - (\nabla \cdot \xi)(x)}{h}$$

is also coordinate-free. □

In the case where $\xi(x) = (\nabla_p H)(x, \nabla u(x)) \iff \nabla u(x) = (\nabla_v L)(x, \xi(x))$ for some potential $u : \mathbb{R}^d \to \mathbb{R}$ (cf. Proposition 3.12, Lemma 4.1), we can derive an explicit formula for $\mathcal{K}_x(\xi)$.

Theorem 5.8 (Formula for $\mathcal{K}_x(\xi)$) *Let $\xi \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ be such that there exists $u \in C^2(\mathbb{R}^d)$, with*

$$\nabla u(x) = (\nabla_v L)(x, \xi(x)), \quad \forall x \in \mathbb{R}^d.$$

Then,

$$\begin{aligned}\mathcal{K}_x(\xi) = {} & L^{ik} \frac{\partial \xi_j}{\partial x_k} \frac{\partial^2 L}{\partial v_j \partial v_l} \frac{\partial \xi_l}{\partial x_i} - L^{im} \frac{\partial^3 L}{\partial v_m \partial v_j \partial v_k} \xi_l \frac{\partial \xi_j}{\partial x_i} \frac{\partial \xi_k}{\partial x_l}\\ & + L^{im} \frac{\partial^3 L}{\partial v_m \partial v_j \partial v_k} L^{kl} \frac{\partial L}{\partial x_l} \frac{\partial \xi_j}{\partial x_i} - L^{ir} \frac{\partial^3 L}{\partial v_r \partial v_j \partial v_k} L^{kl} \frac{\partial^2 L}{\partial x_l \partial v_m} \frac{\partial \xi_j}{\partial x_i} \xi_m\\ & - L^{kl} \frac{\partial^3 L}{\partial x_k \partial v_j \partial v_l} \frac{\partial \xi_j}{\partial x_i} \xi_i + L^{ij} \frac{\partial^3 L}{\partial x_i \partial v_j \partial v_k} L^{kl} \frac{\partial L}{\partial x_l}\\ & - L^{ij} \frac{\partial^3 L}{\partial x_i \partial v_j \partial v_k} L^{kl} \frac{\partial^2 L}{\partial x_l \partial v_m} \xi_m - L^{ij} \frac{\partial^2 L}{\partial x_j \partial x_i}\end{aligned}$$

where all terms involving L are evaluated at $(x, \xi(x))$.

Proof See Appendix. □

In conclusion, the displacement Hessian formula (4.3) can be written as

$$\frac{d^2}{dt^2} \mathcal{F}(\rho) = \int_{\mathbb{R}^d} (\rho G'(\rho) - G(\rho))(\nabla \cdot V)^2 + G(\rho)\mathcal{K}_x(V)\, dx. \tag{5.12}$$

6 Displacement convexity for a non-Riemannian Lagrangian cost

In this section, we provide an example of a Lagrangian cost that is not a squared Riemannian distance. We prove using a perturbation argument that the corresponding generalized curvature is non-negative and thus the generalized entropy functional is convex along C^2 displacement interpolants.

Let $g(x)$ be a positive definite matrix for every $x \in \mathbb{R}^d$ so that $\frac{1}{2}\langle v, g(x)v\rangle$, $v \in \mathbb{R}^d$ defines a Riemannian metric. Let $g_{ij} = g_{ij}(x)$ denote the ij-th entry of $g(x)$ and $g^{ij} = g^{ij}(x)$ denote the ij-th entry of the inverse matrix $g(x)^{-1}$. Further assume that the g_{ij} are bounded with bounded derivatives, and that the corresponding Bakry–Emery tensor (denoted BE_g) is bounded from below. That is,

$$\mathrm{BE}_g = \mathrm{Ric} + \nabla^2\left(\frac{1}{2}\log(\det g)\right) \geq k_g > 0.$$

Define the Lagrangian

$$L(x, v) = \frac{1}{2}\langle v, g(x)v\rangle$$

and the perturbed Lagrangian

$$\tilde{L}(x, v) = \frac{1}{2}\langle v, g(x)v\rangle + \varphi(v)$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a smooth perturbation (for instance, take φ to be of Schwartz class). Using Theorem 5.8, the respective generalized curvatures are given by

$$\begin{aligned}\mathcal{K}_x(\xi) = &\underbrace{g^{ik}\frac{\partial \xi_j}{\partial x_k} g_{jl}\frac{\partial \xi_l}{\partial x_i}}_{I} - \underbrace{g^{kl}\frac{\partial g_{jl}}{\partial x_k}\frac{\partial \xi_j}{\partial x_i}\xi_i}_{V} \\ &+ \underbrace{g^{ij}\frac{\partial g_{jk}}{\partial x_i} g^{kl}\frac{\partial g_{mn}}{\partial x_l}\xi_m\xi_n}_{VI} \\ &- \underbrace{g^{ij}\frac{\partial g_{jk}}{\partial x_i} g^{kl}\frac{\partial g_{nm}}{\partial x_l}\xi_n\xi_m}_{VII} - \underbrace{g^{ij}\frac{\partial^2 g_{kl}}{\partial x_j \partial x_i}\xi_k\xi_l}_{VIII}\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathcal{K}}_x(\xi) = &\underbrace{\tilde{L}^{ik}\frac{\partial \xi_j}{\partial x_k}\frac{\partial^2 \tilde{L}}{\partial v_j \partial v_l}\frac{\partial \xi_l}{\partial x_i}}_{\tilde{I}} - \underbrace{\tilde{L}^{im}\frac{\partial^3 \varphi}{\partial v_m \partial v_j \partial v_k}\xi_l\frac{\partial \xi_j}{\partial x_i}\frac{\partial \xi_k}{\partial x_l}}_{\tilde{II}} \\ &+ \underbrace{\tilde{L}^{im}\frac{\partial^3 \varphi}{\partial v_m \partial v_j \partial v_k}\tilde{L}^{kl}\frac{\partial g_{nr}}{\partial x_l}\xi_n\xi_r\frac{\partial \xi_j}{\partial x_i}}_{\tilde{III}} - \underbrace{\tilde{L}^{ir}\frac{\partial^3 \varphi}{\partial v_r \partial v_j \partial v_k}\tilde{L}^{kl}\frac{\partial g_{mn}}{\partial x_l}\xi_n\frac{\partial \xi_j}{\partial x_i}\xi_m}_{\tilde{IV}} \\ &- \underbrace{\tilde{L}^{kl}\frac{\partial g_{jl}}{\partial x_k}\frac{\partial \xi_j}{\partial x_i}\xi_i}_{\tilde{V}} + \underbrace{\tilde{L}^{ij}\frac{\partial g_{jk}}{\partial x_i}\tilde{L}^{kl}\frac{\partial g_{mn}}{\partial x_l}\xi_m\xi_n}_{\tilde{VI}}\end{aligned}$$

$$- \underbrace{\tilde{L}^{ij} \frac{\partial g_{jk}}{\partial x_i} \tilde{L}^{kl} \frac{\partial g_{mn}}{\partial x_l} \xi_n \xi_m}_{\tilde{VII}} - \underbrace{\tilde{L}^{ij} \frac{\partial^2 g_{kl}}{\partial x_j \partial x_i} \xi_k \xi_l}_{\tilde{VIII}}$$

By Theorem A.3.1 of [19] and (5.7), $\mathcal{K}_x(\xi) = ||g^{-1}\nabla\xi^\top||_{\text{HS}}^2 + \text{BE}_g(\xi)$, where $||\cdot||_{\text{HS}}$ denotes the Hilbert-Schmidt norm. Thus, we have a lower bound

$$\mathcal{K}_x(\xi) = ||g^{-1}\nabla\xi^\top||_{\text{HS}}^2 + \text{BE}_g(\xi) \geq c_g||\nabla\xi||^2 + k_g||\xi||^2$$

where $c_g > 0$ is a constant depending on g. Fix $\epsilon > 0$ such that $\epsilon \leq \min\{\frac{c_g}{10}, \frac{k_g}{12}\}$. Our goal is to choose φ so that

1. $|\tilde{L}^{ij} - L^{ij}| = |\tilde{L}^{ij} - g^{ij}|$ is sufficiently small, i.e. $||\nabla^2\varphi||$ is close to zero, and
2. $|\frac{\partial^3\varphi}{\partial v_i \partial v_j \partial v_k}|$ is sufficiently small.

To this end, we choose φ such that

$$\begin{aligned}
|\tilde{\mathcal{K}}_x(\xi) - \mathcal{K}_x(\xi)| &\leq |\tilde{I} - I| + |\tilde{V} - V| + |\tilde{VI} - VI| + |\tilde{VII} - VII| + |\tilde{VIII} - VIII| \\
&\quad + |\tilde{II}| + |\tilde{III}| + |\tilde{IV}| \\
&\leq \epsilon||\nabla\xi||^2 + 2\epsilon||\nabla\xi||||\xi|| + \epsilon||\xi||^2 + \epsilon||\xi||^2 + \epsilon||\xi||^2 \\
&\quad + \epsilon||\nabla\xi||^2 + 2\epsilon||\nabla\xi||||\xi|| + 2\epsilon||\nabla\xi||||\xi|| \\
&\leq 5\epsilon||\nabla\xi||^2 + 6\epsilon||\xi||^2 \\
&\leq \frac{c_g}{2}||\nabla\xi||^2 + \frac{k_g}{2}||\xi||^2
\end{aligned}$$

Since $\mathcal{K}_x(\xi) \geq c_g||\nabla\xi||^2 + k_g||\xi||^2$, we conclude that $\tilde{\mathcal{K}}_x(\xi) \geq 0$.

Acknowledgements This paper grew out of an undergraduate project I worked on at UCLA. I would like to thank Wilfrid Gangbo (who first introduced this subject to me) for his continued guidance and support. His expertise and generosity have been tremendously helpful. I also thank Tommaso Pacini for helpful feedback. Lastly, I would like to thank the reviewer for the suggestions to the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

7 Appendix

The generalized curvature $\mathcal{K}_x(\xi)$ is given by

$$\mathcal{K}_x(\xi) := \text{tr}\left(\nabla\xi(x)^2 + A(x, \xi(x))\nabla\xi(x) + B(x, \xi(x))\right).$$

In the computations below, time derivatives of ξ will be treated by extending ξ for a short time (in the sense of Proposition 5.3).

Lemma 7.1

$$\mathrm{tr}\left(\nabla\xi(x)^2 + A(x,\xi(x))\nabla\xi(x)\right) = L^{ik}\frac{\partial\xi_j}{\partial x_k}\frac{\partial^2 L}{\partial v_j\partial v_l}\frac{\partial\xi_l}{\partial x_i} - L^{im}\frac{\partial^3 L}{\partial v_m\partial v_j\partial v_k}\xi_l\frac{\partial\xi_j}{\partial x_i}\frac{\partial\xi_k}{\partial x_l}$$
$$+ L^{im}\frac{\partial^3 L}{\partial v_m\partial v_j\partial v_k}L^{kl}\frac{\partial L}{\partial x_l}\frac{\partial\xi_j}{\partial x_i}$$
$$- L^{ir}\frac{\partial^3 L}{\partial v_r\partial v_j\partial v_k}L^{kl}\frac{\partial^2 L}{\partial x_l\partial v_m}\frac{\partial\xi_j}{\partial x_i}\xi_m$$

Proof Recall that

$$A(x,v) = (\nabla^2_{vv}L)(x,v)^{-1}\bigg[\frac{d}{dt}\big((\nabla^2_{vv}L)(\gamma_{x,v}(t),\dot\gamma_{x,v}(t))\big)\Big|_{t=0}$$
$$+ (\nabla^2_{vx}L)(x,v) - (\nabla^2_{xv}L)(x,v)\bigg].$$

By assumption, there exists a potential $u(x)$ satisfying

$$\nabla u(x) = (\nabla_v L)(x,\xi).$$

Since the Hessian

$$\nabla^2 u(x) = (\nabla^2_{vx}L)(x,\xi) + (\nabla^2_{vv}L)(x,\xi)\nabla\xi$$

is symmetric, we have

$$(\nabla^2_{vx}L)(x,\xi) + (\nabla^2_{vv}L)(x,\xi)\nabla\xi = (\nabla^2_{xv}L)(x,\xi) + \nabla\xi^\top(\nabla^2_{vv}L)(x,\xi)$$
$$\implies (\nabla^2_{vv}L)(x,\xi)^{-1}\Big[(\nabla^2_{vx}L)(x,\xi) - (\nabla^2_{xv}L)(x,\xi)\Big] = (\nabla^2_{vv}L)(x,\xi)^{-1}\nabla\xi^\top(\nabla^2_{vv}L)(x,\xi) - \nabla\xi$$

Next,

$$\frac{d}{dt}\big((\nabla^2_{vv}L)(x,\xi)\big)_{ij} = \langle\left(\nabla_v\frac{\partial^2 L}{\partial v_i\partial v_j}\right)(x,\xi),\dot\xi\rangle$$
$$= \langle\left(\nabla_v\frac{\partial^2 L}{\partial v_i\partial v_j}\right)(x,\xi), -\nabla\xi\xi$$
$$+ (\nabla^2_{vv}L)(x,\xi)^{-1}\big[(\nabla_x L)(x,\xi) - (\nabla^2_{xv}L)(x,\xi)\xi\big]\rangle$$
$$= -\frac{\partial^3 L}{\partial v_i\partial v_j\partial v_k}\frac{\partial\xi_k}{\partial x_l}\xi_l + \frac{\partial^3 L}{\partial v_i\partial v_j\partial v_k}L^{kl}\frac{\partial L}{\partial x_l} - \frac{\partial^3 L}{\partial v_i\partial v_j\partial v_k}L^{kl}\frac{\partial^2 L}{\partial x_l\partial v_m}\xi_m$$

□

Lemma 7.2

$$\mathrm{tr}\left(B(x,\xi(x))\right) = -L^{kl}\frac{\partial^3 L}{\partial x_k\partial v_j\partial v_l}\frac{\partial\xi_j}{\partial x_i}\xi_i + L^{ij}\frac{\partial^3 L}{\partial x_i\partial v_j\partial v_k}L^{kl}\frac{\partial L}{\partial x_l}$$
$$- L^{ij}\frac{\partial^3 L}{\partial x_i\partial v_j\partial v_k}L^{kl}\frac{\partial^2 L}{\partial x_l\partial v_m}\xi_m - L^{ij}\frac{\partial^2 L}{\partial x_j\partial x_i}$$

Proof Recall that

$$B(x,v) = (\nabla^2_{vv}L)(x,v)^{-1}\bigg[\frac{d}{dt}\big((\nabla^2_{vx}L)(\gamma_{x,v}(t),\dot\gamma_{x,v}(t))\big)\Big|_{t=0} - (\nabla^2_{xx}L)(x,v)\bigg].$$

By a similar computation as the previous lemma, we have

$$\begin{aligned}\frac{d}{dt}\big((\nabla^2_{vx}L)(x,\xi)\big)_{ij} &= \langle\left(\nabla_v \frac{\partial^2 L}{\partial v_i \partial x_j}\right)(x,\xi), \dot{\xi}\rangle \\ &= \langle\left(\nabla_v \frac{\partial^2 L}{\partial v_i \partial x_j}\right)(x,\xi), -\nabla\xi\xi \\ &\quad + (\nabla^2_{vv}L)(x,\xi)^{-1}\big[(\nabla_x L)(x,\xi) - (\nabla^2_{xv}L)(x,\xi)\xi\big]\rangle\end{aligned}$$

□

Putting together these two lemmas, we get the formula for $\mathcal{K}_x(\xi)$.

References

1. Agueh, M.: Sharp Gagliardo–Nirenberg inequalities and mass transport theory. J. Dyn. Differ. Equ. **18**, 1069–1093 (2006)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhäuser, New York (2005)
3. Bauer, M., Modin, K.: Semi-invariant Riemannian metrics in hydrodynamics. Calc. Var. Partial. Differ. Equ. **59**(2), Paper No. 65, 25 pp. (2020)
4. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. **84**, 375–393 (2000)
5. Bernard, P., Buffoni, B.: Optimal mass transportation and Mather theory. J. Eur. Math. Soc. **9**, 85–121 (2007)
6. Cannarsa, P., Sinestrari, C.: Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control. Birkhäuser, New York (2004)
7. Carrillo, J.A., Slepčev, D.: Example of a displacement convex functional of first order. Calc. Var. Partial. Differ. Equ. **36**, 547–564 (2008)
8. Cordero-Erausquin, D., Gangbo, W., Houdré, C.: Inequalities for generalized entropy and optimal transportation. Contemp. Math. **353**, 73–94 (2004)
9. Fathi, A.: Weak KAM theorem in Lagrangian dynamics. Monograph **88** (2014)
10. Fathi, A.: Viscosity solutions of the Hamilton–Jacobi equation on a non-compact manifold (2020)
11. Fathi, A., Figalli, A.: Optimal transportation on non-compact manifolds. Israel J. Math. **175**, 1–59 (2007)
12. Figalli, A., Gangbo, W., Yolcu, T.: A variational method for a class of parabolic PDEs. Annali della Scuola Normale Superiore di Pisa - Classe di Scienze **Ser. 5, 10**(1), 207–252 (2011)
13. Figalli, A., Gigli, N.: Local semiconvexity of Kantorovich potentials on non-compact manifolds, ESAIM: control. Optim. Cal. Var. **17**(3), 648–653 (2011)
14. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math. **177**(2), 113–161 (1996)
15. Gomes, D.A., Seneci, T.: Displacement convexity for first-order mean-field games. Minimax Theory Appl. **3**(2), 261–284 (2018)
16. McCann, R.J.: A convexity principle for interacting gases. Adv. Math. **128**(1), 153–179 (1997)
17. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Part. Differ. Equ. **26**, 101–174 (2001)
18. 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)
19. Schachter, B.: An Eulerian Approach to Optimal Transport with Applications to the Otto Calculus. Ph.D. thesis (2017)
20. Schachter, B.: A new class of first order displacement convex functionals. SIAM J. Math. Anal. **50**(2), 1779–1789 (2018)
21. Villani, C.: Optimal transport—old and new, Grundlehren Math. Wiss, vol. 338. Springer, Berlin (2008)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.