OXFORD

Simple Unbalanced Optimal Transport

Boris Khesin[1], Klas Modin[2,*] and Luke Volk[1]

[1]Department of Mathematics, University of Toronto, Ontario M5S 2E4, Canada
[2]Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden
*Correspondence to be sent to: e-mail: klas.modin@chalmers.se
Communicated by Prof. Justin Solomon

We introduce and study a simple model capturing the main features of unbalanced optimal transport. It is based on equipping the conical extension of the group of all diffeomorphisms with a natural metric, which allows a Riemannian submersion to the space of volume forms of arbitrary total mass. We describe its finite-dimensional version and present a concise comparison study of the geometry, Hamiltonian features, and geodesics for this and other extensions. One of the corollaries of this approach is that along any geodesic the total mass evolves with constant acceleration, as an object's height in a constant buoyancy field.

1 Introduction

Many problems of optimal transport are closely related to the differential geometry of diffeomorphism groups. In particular, the problem of moving one mass (or density) to another by a diffeomorphism while minimizing a certain (quadratic) cost can be understood as construction of geodesics in an appropriate metric on the space of normalized densities (or on its completion); see, for example, [17, 23]. Similar problems arise in applications when one attempts to evaluate the proximity between different shapes or medical images [21]. However, the action by a diffeomorphism does not allow a change of the total mass of the density. Hence, one arrives at the problem of constructing a natural extension of the action which would allow one to connect in the most economical way densities of different total masses. Such problems, first considered by Benamou [2], belong to the domain of *unbalanced optimal transport* (UOT), and they have received considerable attention lately; see [5, 6, 11, 12, 18, 22] for geometry and analysis and [1, 4, 19] for numerical aspects.

Usually, the setting of unbalanced optimal transport involves a "large" extension $G = \mathrm{Diff}(M) \ltimes C_+^\infty(M)$ of the group $\mathrm{Diff}(M)$ of all diffeomorphisms of a manifold by means of a semi-direct product with the space of smooth positive functions. Such a large semi-direct product group acts on densities by a change of coordinates and then by adjusting point-wise the obtained density by means of a function.

In this paper, we instead introduce and study a much simpler "small" extension $\mathbf{cone}(\mathrm{Diff}(M)) = \mathrm{Diff}(M) \times \mathbb{R}_+$ of the same group $\mathrm{Diff}(M)$. This way, both the group of diffeomorphisms and the space of normalized densities have similar conical extensions $\mathbf{cone}(\mathrm{Diff}(M))$ and $\mathrm{Vol}(M) = \mathbf{cone}(\mathrm{Dens}(M))$ by one extra parameter, the total mass m of the density. We describe natural metrics and geodesics for those extensions.

It turns out that the corresponding problem of unbalanced optimal transport, while being much easier to handle, captures most of the main features for the large extensions. For instance, for both small and large extensions, a common phenomenon is that in many two-point problems a geodesic

joining two end-densities goes through densities whose total mass dips below the smallest of the two it connects. In particular, one of the corollaries of this approach is that along any geodesic the total mass m evolves with constant acceleration, $\ddot{m} = const$, that is, as an object's height in a constant buoyancy field.

We also introduce special variables in which we demonstrate the convexity of the dynamical formulation for the simple conical extension, generalizing the convexity of standard optimal transport. This convex minimization formulation is known to be central for the existence and uniqueness of the corresponding solutions in such variational problems.

One immediate additional advantage of the present approach is that it admits a finite-dimensional model, where the diffeomorphism group Diff(M) for $M = \mathbb{R}^n$ is replaced by its subgroup $\mathbf{GL}(n)$, while the space of all volume forms is constrained to its subspace of non-normalized Gaussian densities on $\mathbb{R}^n$.

Finally, we compare in more detail our "small" extension with two other "larger" extensions: the one considered in [5, 6] and called Wasserstein-Fisher-Rao and the one which is indeed a weighted sum of the Wasserstein and Fisher-Rao metrics. In a sense, those two models can be viewed as extensions of our simpler model in, respectively, Lagrangian and Hamiltonian settings, as we discuss below. We describe the corresponding geodesics and candidates for their finite-dimensional counterparts. It turns out that the corresponding larger finite-dimensional models are less natural than for the small extension, as they require additional restrictions on orbits of the corresponding action.

2   A Conical Extension of the Diffeomorphism Group

Let M be an n-dimensional Riemannian manifold with volume form μ of total volume (or "mass") equal to 1. Let Vol(M) denote the set of all (un-normalized) volume forms on M of finite total volume. While for most applications one can think of a compact manifold M, it is also convenient to keep in mind the case of $M = \mathbb{R}^n$ with Gaussian densities on it.

Throughout the paper, we consider infinite-dimensional manifolds and Lie groups, such as the spaces of smooth normalized densities **Dens**(M), smooth volume forms Vol(M), and smooth diffeomorphisms Diff(M). In the smooth, C^∞ category, the manifold structures are modelled on Fréchet spaces. Alternatively, one can work in the category of Banach manifold via completions in the Sobolev H^s category (which requires $s > n/2 + 1$ as a consequence of the Sobolev embedding theorem). For details on these settings, we refer to [8, 9, 15] and references therein.

Let Diff(M) $\times \mathbb{R}_+$ denote the direct product Lie group. A left action of Diff(M) $\times \mathbb{R}_+$ on Vol(M) is given by $(\varphi, m) \cdot \varrho = m\varphi_* \varrho$ for $\varphi \in$ Diff(M), $\varrho \in$ Vol(M), and $m \in \mathbb{R}_+$. Fixing a Riemannian volume form μ, this left action endows the product Diff(M) $\times \mathbb{R}_+$ with the structure of a principal G-bundle with projection

$$\pi : \mathrm{Diff}(M) \times \mathbb{R}_+ \to \mathrm{Vol}(M)$$

$$(\varphi, m) \mapsto m\,\varphi_* \mu$$

and corresponding isotropy subgroup G given by

$$G = \{(\varphi, m) \mid m\varphi_* \mu = \mu\} = \mathrm{Diff}_\mu(M) \times \{1\}.$$
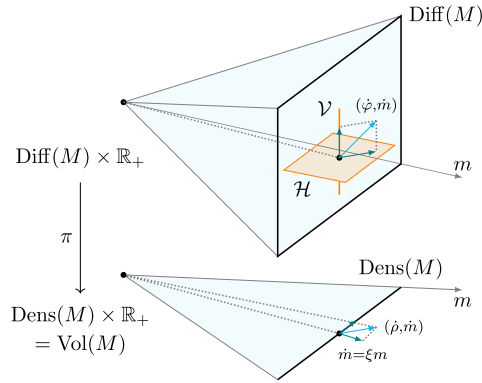
It follows that $m = 1$ by taking the integral. The Lie algebra of G is thus

$$\mathfrak{g} = \mathfrak{X}_\mu(M) \times \{0\}.$$

The tangent space of the fibre through (φ, m), denoted $\mathcal{V}_{(\varphi, m)} = \ker d\pi_{(\varphi, m)}$, gives the vertical distribution associated with the bundle π.

Lemma 2.1. The vertical distribution for $\pi : \mathrm{Diff}(M) \times \mathbb{R}_+ \to \mathrm{Vol}(M)$ is given by

$$\mathcal{V}_{(\varphi, m)} = \left\{ (v \circ \varphi, 0) \mid \mathbf{div}(\rho v) = 0 \quad \text{for} \quad \rho = \frac{m\,\varphi_* \mu}{\mu} \right\}.$$

Fig. 1. Illustration of the submersion between two conical extensions.

Proof. Given a curve $(\varphi(t), m(t)) \in \mathrm{Diff}(M) \times \mathbb{R}_+$ with $(\varphi(0), m(0)) = (\varphi, m)$, we have that for $(\dot{\varphi}(0), \dot{m}(0)) = (\upsilon \circ \varphi, \xi\, m) \in T_{(\varphi,m)}(\mathrm{Diff}(M) \times \mathbb{R}_+)$:

$$d\pi_{(\varphi,m)}(\upsilon \circ \varphi, \xi m) = \left.\frac{d}{dt}\right|_{t=0} m(t)\varphi(t)_*\mu$$

$$= \xi m\, \varphi_*\mu + m\, \left.\frac{d}{dt}\right|_{t=0} \varphi(t)_*\mu \tag{1}$$

$$= \xi\rho\mu - m\, L_\upsilon \varphi_*\mu = \xi\rho\mu - m\, \mathbf{div}(\rho\upsilon)\mu.$$

Thus, $(\upsilon \circ \varphi, \xi m) \in \mathcal{V}_{(\varphi,m)}$ if and only if $\xi\rho = m\,\mathbf{div}(\rho\upsilon)$. Now, by integrating the both sides against μ over M we see that the integral of the divergence is zero. This implies that the constant $\xi = 0$, which in turn implies that the divergence is zero point-wise. This concludes the proof. ∎

2.1 A natural metric for UOT

Consider the following metric on the direct product group $\mathrm{Diff}(M) \times \mathbb{R}_+$:

$$\mathcal{G}_{(\varphi,m)}((\dot{\varphi}, \dot{m}), (\dot{\varphi}, \dot{m})) = m \int_M |\dot{\varphi}|^2\mu + \frac{\dot{m}^2}{m} = \int_M |\upsilon|^2\varrho + m\xi^2 \tag{2}$$

for variables $\upsilon = \dot{\varphi} \circ \varphi^{-1}$, $\xi = \dot{m}/m$, and $\varrho = m\varphi_*\mu$.

Remark 2.2. Recall that for a Riemannian manifold N with metric $g(\upsilon, \upsilon)$ its conical extension **cone**$(N) := N \times \mathbb{R}_+$ is a Riemannian manifold with metric $r^2 g(\upsilon, \upsilon) + dr^2$. Consequently, the above product group $\mathrm{Diff}(M) \times \mathbb{R}_+$ is a natural conical extension of the most straightforward L^2 metric on $\mathrm{Diff}(M)$ given by

$$\langle \dot{\varphi}, \dot{\varphi} \rangle_\varphi = \int_M |\dot{\varphi}|^2\mu\,.$$

Indeed, by changing variables $m = r^2$ (implying $\dot{m} = 2r\dot{r}$) we come to the conical extension $\mathrm{Diff}(M) \times \mathbb{R}_+$ with metric

$$\langle (\dot{\varphi}, \dot{r}), (\dot{\varphi}, \dot{r}) \rangle_{(\varphi,r)} = r^2 \int_M |\dot{\varphi}|^2\mu + 4\dot{r}^2\,. \tag{3}$$

The orthogonal complement of the vertical distribution with respect to the metric on $\mathrm{Diff}(M) \times \mathbb{R}_+$ gives the horizontal distribution of the bundle.

Lemma 2.3. For the metric $\mathcal{G}$ in equation (2), the horizontal distribution at (φ, m) is given by

$$\mathcal{H}_{(\varphi,m)} = \{(\nabla\theta \circ \varphi, \xi m) \mid \theta \in C^\infty(M), \xi \in \mathbb{R}\} \simeq \{\theta \in C^\infty(M)\}\,,$$

$$\text{where} \quad \left(\nabla\theta \circ \varphi, \int_M \theta\varrho\right) \leftrightarrow \theta\,.$$

Proof. The vertical distribution $\mathcal{V}_{(\varphi,m)}$ consists of $(\upsilon \circ \varphi, 0)$ where υ is divergence free with respect to $\varrho = m\varphi_* \mu$, that is, $\mathbf{div}(\rho\upsilon) = 0$. Thus, it follows from the (generalized) Hodge decomposition and the choice of metric (2) that if $(u \circ \varphi, \xi m) \in \mathcal{H}$ then $u = \nabla\theta$ is a gradient vector field. It now follows that $(\nabla\theta \circ \varphi, \xi m)$ is orthogonal to $\mathcal{V}_{(\varphi,m)}$ for any $\xi \in \mathbb{R}$. In particular, we may encode ξ in the arbitrary constant of θ for $\nabla\theta$. The choice $\xi m = \int_M \theta \varrho$ gives a geometric identification of $\mathcal{H}_{(\varphi,m)}$ with the space $C^\infty(M)$. ∎

Theorem 2.4. The metric $\mathcal{G}$ in (2) projects as a Riemannian submersion to the metric $\bar{\mathcal{G}}$ on Vol(M) given at any point $\rho \in \text{Vol}(M)$ by

$$\bar{\mathcal{G}}_\varrho(\dot{\varrho}, \dot{\varrho}) = \int_M \left(|\nabla\theta|^2 + \xi^2 \right) \varrho, \quad \dot{\rho} = -\mathbf{div}(\rho\nabla\theta) + \xi\rho, \quad \int_M \dot{\varrho} = m\xi. \tag{4}$$

Furthermore, the variable $\theta \in C^\infty(M)$, defined by the equations above together with

$$\xi m = \int_M \theta\varrho,$$

is Legendre-dual to $\dot{\varrho}$ under the pairing

$$\langle \dot{\varrho}, \theta \rangle = \int_M \theta\dot{\varrho}.$$

Consequently, the Hamiltonian on T*Vol(M) corresponding to the metric $\bar{\mathcal{G}}$ is

$$H(\varrho, \theta) = \frac{1}{2} \int_M |\nabla\theta|^2 \varrho + \underbrace{\frac{1}{2m} \left(\int_M \theta\varrho \right)^2}_{\frac{1}{2}m\xi^2}. \tag{5}$$

Proof. First, notice that the metric $\mathcal{G}$ is invariant under the right action of the isotropy subgroup G on the tangent bundle $T(\mathbf{Diff}(M) \times \mathbb{R}_+)$. Thus, $\mathcal{G}$ is compatible with the principal bundle structure, so it indeed induces a metric $\bar{\mathcal{G}}$ on the base $\mathbf{Vol}(M)$. Now take an arbitrary horizontal vector $(\nabla\theta \circ \varphi, \xi m) \in \mathcal{H}_{(\varphi,m)}$. If $\varrho = \pi(\varphi, m)$ and $\dot{\varrho} = d\pi_{(\varphi,m)}(\nabla\theta \circ \varphi, \xi m)$ is the lifted bundle projection, then, by definition,

$$\bar{\mathcal{G}}_\varrho(\dot{\varrho}, \dot{\varrho}) \equiv \mathcal{G}_{(\varphi,m)}(\nabla\theta \circ \varphi, \xi m, \nabla\theta \circ \varphi, \xi m) = \int_M |\nabla\theta|^2 \varrho + m\xi^2.$$

From equation (1) for $d\pi$ we get that $\dot{\rho} = -\mathbf{div}(\rho\nabla\theta) + \xi\rho$. Applying integration and using that $m = \int_M \varrho$, we see that

$$\int_M \dot{\varrho} = m\xi.$$

This confirms the formula (4) for the induced metric.

For the second statement, that θ is in fact the Legendre transform, the variable Legendre-dual to $\dot{\varrho}$ is defined by $\frac{\delta L}{\delta \dot{\varrho}}$ where L is the Lagrangian corresponding to $\bar{\mathcal{G}}$. Given a variation $\dot{\varrho}_\epsilon = \dot{\varrho} + \epsilon \, \delta\dot{\varrho}$ we obtain

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} L(\varrho, \dot{\varrho}_\epsilon) = \underbrace{\int_M \left(\nabla\theta \cdot \nabla \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \theta_\epsilon \right) \varrho}_{(i)} + \underbrace{\xi m \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \xi_\epsilon}_{(ii)}. \tag{6}$$

On the other hand, from the definition of θ in (4) we see that

$$\delta\dot{\rho} = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \dot{\rho}_\epsilon = -\mathbf{div}\left(\rho\nabla \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \theta_\epsilon \right) + \rho \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \xi_\epsilon. \tag{7}$$

By applying the divergence theorem to the term (i) and then comparing (6) with (7), we see that $\langle \delta\dot{\varrho}, \theta \rangle = \frac{d}{d\epsilon}\big|_{\epsilon=0} L(\varrho, \dot{\varrho}_\epsilon)$, giving θ as the Legendre-dual variable of $\dot{\varrho}$. The form of the Hamiltonian follows readily. ∎

Equipping Diff(M) $\times \mathbb{R}_+$ and Vol(M) with the metrics $\mathcal{G}$ and $\bar{\mathcal{G}}$ (see (2) and (4)) makes $\pi : \text{Diff}(M) \times \mathbb{R}_+ \to$ Vol(M) into a Riemannian submersion, which gives a correspondence between geodesics in Vol(M) and

horizontal geodesics in $\mathrm{Diff}(M) \times \mathbb{R}_+$ (i.e., those tangent to the horizontal distribution), when given an initial point in the fiber.

2.2 Dynamical and static formulations

One can give the following dynamical formulation of conical unbalanced transport.

Definition 2.5. The *conical Wasserstein distance* $\mathrm{WC}(\varrho_0, \varrho_1)$ between densities $\varrho_0, \varrho_1 \in \mathrm{Vol}(M)$ (of possibly different total masses) is given by the following formula:

$$\mathrm{WC}^2(\varrho_0, \varrho_1) = \inf_{u, \xi, \varrho} \int_0^1 \left(\int_M \left(|u|^2 + \xi^2 \right) \varrho \right) \, dt,$$

over time-dependent vector fields u, volume forms ϱ, and constants ξ related by the constraints

$$\dot{\rho} = -\operatorname{div}(\rho u) + \xi \rho, \quad \int_M \dot{\varrho} = \xi \int_M \varrho, \quad \rho(0) = \rho_0, \quad \rho(1) = \rho_1.$$

The convexity of the dynamical formulation of standard optimal transport, as studied by Benamou and Brenier [3], carries over to the conical extension. Indeed, in the variables $\bar{\rho} = \rho/m = \frac{\varphi_* \mu}{\mu} > 0$, $w = \bar{\rho} \nabla \theta$, and $r = \sqrt{m} > 0$ it becomes

$$\mathrm{WC}^2(\varrho_0, \varrho_1) = \inf_{w, \bar{\rho}, r} \int_0^1 \left(r^2 \int_M \frac{|w|^2}{\bar{\rho}} \mu + 4\dot{r}^2 \right) \, dt,$$

that is, a minimization of a convex functional, under the affine constraints

$$\dot{\bar{\rho}} + \operatorname{div} w = 0, \quad \bar{\rho}(0, \cdot) = \rho_0/m_0, \quad \bar{\rho}(1, \cdot) = \rho_1/m_1,$$
$$r(0) = \sqrt{m_0}, \quad r(1) = \sqrt{m_1}.$$

This convex minimization formulation is important for existence and uniqueness of solutions.

For the corresponding static formulation, the distance function $\mathrm{WC}(\varrho_0, \varrho_1)$ is given in terms of the Riemannian metric (4) as

$$\mathrm{WC}^2(\varrho_0, \varrho_1) = \inf_{\varrho} \int_0^1 \bar{\mathcal{G}}_\varrho(\dot{\varrho}, \dot{\varrho}) \, dt,$$

for curves $\varrho(t)$ with $\varrho(0) = \varrho_0$ and $\varrho(1) = \varrho_1$.

Note that the distance function WC is necessarily implicit, as it depends on the metric on the Riemannian manifold M. It is bounded above via the (adjusted) Wasserstein distance W between denities of the unit total mass as follows: for densities ϱ_0 and ϱ_1 of masses m_0 and m_1, respectively, one has the upper bound

$$\mathrm{WC}^2(\varrho_0, \varrho_1) \leq \min(m_0, m_1) \cdot W^2(\varrho_0/m_0, \varrho_1/m_1) + 4(\sqrt{m_1} - \sqrt{m_0})^2 \, .$$

It follows from the orthogonality of the radial direction $r = \sqrt{m}$ to $\mathrm{Dens}(M)$.

2.3 Geodesic equations

The equations of geodesics for the above metrics can be computed in either Lagrangian or Hamiltonian form. The Lagrangian form of the geodesic equations, that is, equations in the corresponding tangent bundle, for the conical manifold can be obtained using the formulas for warped Riemannian manifolds (see [16]). We review this approach in Appendix A. Here we derive the geodesic equations on the cotangent bundle, that is, as the Hamiltonian equations for the Hamiltonian (5).

Theorem 2.6. The geodesic equations in Hamiltonian form for the Hamiltonian (5) are given by

$$\dot\rho = -\operatorname{div}(\rho\nabla\theta) + \xi\rho$$

$$\dot\theta = -\frac{1}{2}|\nabla\theta|^2 - \xi\theta + \frac{\xi^2}{2}.$$

Proof. Hamilton's equations are $\dot\varrho = \delta H/\delta\theta$ and $\dot\theta = -\delta H/\delta\varrho$. First, consider a variation $\theta_\epsilon = \theta + \epsilon\,\delta\theta$, where:

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} H(\varrho,\theta_\epsilon) = \int_M (\rho\nabla\theta)\cdot\nabla(\delta\theta)\,\mu + \frac{1}{m}\left(\int_M \theta\,\varrho\right)\left(\int_M \delta\theta\,\varrho\right)$$

$$= \int_M (\rho\nabla\theta)\cdot\nabla(\delta\theta)\,\mu + \int_M (\xi\rho)\delta\theta\,\mu$$

$$= \int_M \left(\operatorname{div}((\rho\nabla\theta)\delta\theta) - \operatorname{div}(\rho\nabla\theta)\delta\theta + (\xi\rho)\delta\theta\right)\mu$$

$$= \int_M \left(-\operatorname{div}(\rho\nabla\theta) + \xi\rho\right)\delta\theta\,\mu,$$

and so $\dot\rho = -\operatorname{div}(\rho\nabla\theta) + \xi\rho$.

Similarly, considering a variation $\varrho_\epsilon = \varrho + \epsilon\,\delta\varrho$:

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} H(\varrho_\epsilon,\theta) = \frac{1}{2}\int_M |\nabla\theta|^2\,\delta\varrho - \frac{\dot m_0}{2m^2}\left(\int \theta\,\varrho\right)^2 + \frac{1}{m}\left(\int_M \theta\,\varrho\right)\left(\int_M \theta\,\delta\varrho\right)$$

$$= \frac{1}{2}\int_M |\nabla\theta|^2\,\delta\varrho - \frac{\dot m_0}{2m^2}(m\xi)^2 + \frac{1}{m}(m\xi)\left(\int_M \theta\,\delta\varrho\right),$$

but $\dot m_0 = \frac{d}{d\epsilon}\Big|_{\epsilon=0}\int_M \varrho_\epsilon = \int_M \delta\varrho$, so:

$$= \int_M \left(\frac{1}{2}|\nabla\theta|^2 - \frac{\xi^2}{2} + \xi\theta\right)\delta\varrho,$$

hence $\dot\theta = -\frac{1}{2}|\nabla\theta|^2 - \xi\theta + \frac{\xi^2}{2}$. ∎

Recall from above that $m = \int_M \varrho$ is the total volume and that $\xi = \int_M \theta\varrho/m$ is the logarithmic derivative of m. The evolution of m and ξ is described by the following theorem.

Theorem 2.7. The variables m and ξ satisfy the equations

$$\dot\xi = \frac{1}{m}\left(H(\varrho,\theta) - m\xi^2\right)$$

$$\dot m = m\xi.$$

It can also be written as the second order equation

$$\ddot m = H.$$

Since $H(\varrho,\theta)$ is constant along solutions, we obtain the following:

Corollary 2.8. The total volume $m := \int_M \varrho$ evolves with constant acceleration that depends only on the energy level of the initial conditions.

In other words, the volume m evolves as an object's height in a constant gravity or buoyancy field. Note that in a conical metric it is a common phenomenon that, depending of the boundary conditions, a

geodesic joining two densities might enter the region where the total mass is smaller than the smallest of the two it connects.

Proof. By construction $\dot{m} = \xi m$. From $\xi m = \int_M \theta \varrho$ we then get

$$
\begin{aligned}
\dot{\xi} &= \frac{d}{dt} \frac{1}{m} \int_M \theta \varrho = \frac{1}{m} \int_M \left(\dot{\theta} \varrho + \theta \dot{\varrho} \right) - \xi^2 \\
&= \frac{1}{m} \int_M \left(\left(-\frac{1}{2} |\nabla \theta|^2 - \xi \theta + \frac{\xi^2}{2} \right) \varrho + \theta \left(-\operatorname{div}(\rho \nabla \theta) + \xi \varrho \right) \right) - \xi^2 \\
&= \frac{1}{m} \int_M \left(\frac{1}{2} |\nabla \theta|^2 \right) \varrho - \frac{\xi^2}{2} = \frac{1}{m} \left(\int_M \frac{1}{2} |\nabla \theta|^2 \varrho + \frac{1}{2} m \xi^2 - \frac{1}{2} m \xi^2 \right) - \frac{\xi^2}{2} \\
&= \frac{H(\varrho, \theta)}{m} - \xi^2 .
\end{aligned}
$$

We then obtain that $\ddot{m} = \dot{m}\xi + m\dot{\xi} = m\xi^2 + m\frac{1}{m}(H - m\xi^2) = H.$ ■

Remark 2.9. Note from equation (5) that $H \geq \frac{1}{2} m \xi^2$ with equality if and only if θ is constant, which corresponds to the invariant subset of pure scalings of the density ϱ.

Remark 2.10. In the simple model just presented, mass is added or removed proportionally to ρ. It is easy to modify the model, so it has a localized "production function" $f = f(x) \geq 0$ representing a fixed rate of supply or demand distribution over M. In the model above such a rate was constant, $f \equiv 1$, manifesting that the volume was added or subtracted uniformly over M, while using a nonconstant f one can adjust the UOT model and make some regions of M preferred to others. Then, instead of the Hamiltonian (5) we use

$$
H(\varrho, \theta) = \frac{1}{2} \int_M |\nabla \theta|^2 \varrho + \frac{1}{2m} \left(\int_M \theta f \varrho \right)^2 , \tag{8}
$$

where now $\xi m = \int_M \theta f \varrho$. The evolution of ρ and θ then becomes

$$
\dot{\rho} = -\operatorname{div}(\rho \nabla \theta) + \xi f \rho , \qquad \dot{\theta} = -\frac{1}{2} |\nabla \theta|^2 - \xi f \theta + \frac{\xi^2}{2} .
$$

Remark 2.11. Note that the geodesic equation in Theorem 2.6 retains a property of conical extensions; the radial projection of a geodesic curve in Vol(M) corresponds to a Wasserstein geodesic on the space **Dens**(M) of normalized densities, albeit in a different parameterization and with a different total length. Indeed, it follows from the fact that a totally geodesic submanifold of a manifold remains totally geodesic after its conical extension, see the next lemma. Note, however, that this projection property for geodesics does not hold for extensions with non-constant production functions, cf. Remark 2.10.

Remark 2.12. Another model for unbalanced optimal transport is given in [7]. It is also an extension by means of one extra dimension, and it can be viewed as a *cylindrical*-type, rather than *conical*, extension of the Wasserstein geometry discussed above. The dynamics of density is given by the equation $\dot{\rho} = -\operatorname{div}(\rho u) + h$ where the last term $h = h(t)$ can be regarded as "pumping" the constant density over the whole of manifold, and it replaces the linear term $\xi \rho$ proportional to the current density ρ in the conical model. Then the dynamics of this extra variable $h(t)$ is governed by the vector field u fulfilling the inviscid Burgers equation, as in the standard optimal transport. This leads to the uniform change of $h(t)$ ($\dot{h} = \mathbf{const}$) and a somewhat peculiar numerical behavior observed in [7].

2.4 A finite-dimensional version of the simple UOT

The existence of a finite-dimensional version of the conical extension is based on the following observation.

Lemma 2.13. Suppose that a submanifold $N \subset M$ is a totally geodesic in the manifold M. Then **cone**(N) is totally geodesic in **cone**(M).

Proof. To prove the totally geodesic property, one needs to compute the geodesic equations. One can see that if the covariant derivatives $\nabla_{\dot q} \dot q$ for $q \in N \subset M$ belong to the tangent bundle of N then its extension by the radial variable $r \in \mathbb{R}_+$ can belong to the product of the tangent bundle of N and $\mathbb{R}_+$. ∎

Corollary 2.14. Conical extensions $\mathbf{GL}(n) \times \mathbb{R}_+ \subset \mathrm{Diff}(\mathbb{R}^n) \times \mathbb{R}_+$ of the sub-manifolds $\mathbf{GL}(n) \subset \mathrm{Diff}(\mathbb{R}^n)$ are totally geodesic for the natural UOT metric. (The same statement holds for the unbalanced $\dot H^1$ and Fisher-Rao metrics considered below.)

On the base, we now restrict the metric to the space of scaled (or non-normalized) Gaussian densities $\mathcal{N} \subset \mathrm{Vol}(M)$. In the total space, we restrict the metric to the finite-dimensional direct product subgroup $\mathbf{GL}(n) \times \mathbb{R}_+ \subset \mathrm{Diff}(M) \times \mathbb{R}_+$:

$$\mathcal{G}^{\mathrm{fin}}_{(A,m)}((\dot A, \dot m), (\dot A, \dot m)) = m \int_{\mathbb{R}^n} \|\dot A x\|^2 \eta(x) + \frac{\dot m^2}{m}, \tag{9}$$

where $\varphi(x) = Ax$ for $A \in \mathbf{GL}(n)$ and $\eta = p(x, \Sigma)\, dx$ is a normal density with covariance matrix Σ and zero mean,

$$p(x, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} x^\top \Sigma^{-1} x\right).$$

Remark 2.15. Recall that the isotropic Gaussian is given by

$$\mu(x) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{1}{2} x^\top x\right) dx.$$

Consider now a group element $(\varphi : x \mapsto Ax, m)$. The action on μ is

$$m\, \varphi_* \mu = \sqrt{\frac{m^2}{\det(AA^\top)(2\pi)^n}} \exp\left(-\frac{1}{2} x^\top (AA^\top)^{-1} x\right) dx =: m\, p(x, \underbrace{AA^\top}_{\Sigma}) dx.$$

The latter has the natural scaling property:

$$m\, p(x, \Sigma) = p\left(\frac{x}{\sqrt[n]{m}}, \frac{\Sigma}{\sqrt[n]{m^2}}\right).$$

Note that one cannot write $m\, p(x, \Sigma) = p(x, \tilde\Sigma)$ for some covariance matrix $\tilde\Sigma$, since $p(\cdot, \Sigma_1) = p(\cdot, \Sigma_2) \iff \Sigma_1 = \Sigma_2$.

After identifying the Gaussian densities with their (symmetric positive definite) covariance matrices in $\mathbf{Sym}_+(n)$, the finite-dimensional version of our bundle is

$$\pi : \mathbf{GL}(n) \times \mathbb{R}_+ \to \mathbf{Sym}_+(n) \times \mathbb{R}_+$$

$$(A, m) \mapsto (A\Sigma A^\top, m),$$

where we parametrize the base using both the covariance matrix and the total volume. The metric (9) on $\mathbf{GL}(n) \times \mathbb{R}_+$ in terms of $(A, V) \in T_A \mathbf{GL}(n)$, where $V = \dot A A^{-1}$, and $(m, \xi) \in T_m \mathbb{R}_+$ is given by

$$\mathcal{G}^{\mathrm{fin}}_{(A,m)}((\dot A, \dot m), (\dot A, \dot m)) = m \left(\int_{\mathbb{R}^n} |Vx|^2 p(x, \Sigma) dx + \xi^2 \right)$$

$$= m \left(\mathrm{tr}(\Sigma V^\top V) + \xi^2 \right).$$

Lemma 2.16. The vertical and horizontal distributions of $\mathbf{GL}(n) \times \mathbb{R}_+$ with the metric $\mathcal{G}^{\mathrm{fin}}$ are

$$\mathcal{V}_{(A,m)} = \{(VA, 0) \in T_A \, \mathbf{GL}(n) \times \mathbb{R} \mid 0 = V(A\Sigma A^\top) + (A\Sigma A^\top)V^\top\},$$

$$\mathcal{H}_{(A,m)} = \{(VA, \xi m) \in T_A \, \mathbf{GL}(n) \times \mathbb{R} \mid V \in \mathbf{Sym}(n), \quad \xi \in \mathbb{R}\}.$$

Proof. If $(A(t), m(t))$ is a path in $\mathbf{GL}(n) \times \mathbb{R}_+$ with $(A(0), m(0)) = (A, m)$ and $(\dot{A}(0), \dot{m}(0)) = (VA, \xi m)$, then:

$$d\pi_{(a,m)}(VA, \xi m) = \frac{d}{dt}\bigg|_{t=0} (A(t)\Sigma A(t)^\top, m(t))$$

$$= (\dot{A}(0)\Sigma A(0)^\top + A(0)\Sigma \dot{A}(0)^\top, \dot{m}(0))$$

$$= (VA\Sigma A^\top + A\Sigma A^\top V^\top, \xi m),$$

which gives the desired vertical distribution as its kernel. Noting that $\mathcal{V}_{(A,m)}$ consists of VA such that $VA\Sigma A^\top$ is antisymmetric, if $WA \in \mathcal{H}_{(A,m)}$ then for all such $Z = VA$ we have:

$$0 = \mathcal{G}^{\mathrm{fin}}_{(A,m)}((W, \xi m), (Z, 0)) = m \, \mathrm{tr}(WA\Sigma Z^\top) = -m \, \mathrm{tr}\left(W(Z\Sigma A^\top)\right).$$

Picking $Z\Sigma A^\top$ to be the elementary antisymmetric matrix with 1 in the (i, j)-entry and -1 in the (j, i)-entry (for $i \neq j$) gives that W must be symmetric, giving the desired horizontal distribution. ∎

The projection $\pi \colon \mathbf{GL}(n) \times \mathbb{R}_+ \to \mathbf{Sym}_+(n) \times \mathbb{R}_+$ subduces a metric $\bar{\mathcal{G}}^{\mathrm{fin}}$ on $\mathbf{Sym}_+(n) \times \mathbb{R}_+$ by defining

$$\bar{\mathcal{G}}^{\mathrm{fin}}_{\pi(A,m)}(d\pi_{(A,m)}(X, a), d\pi_{(A,m)}(Y, b)) = \mathcal{G}^{\mathrm{fin}}_{(A,m)}((X, a)_{\mathcal{H}}, (Y, b)_{\mathcal{H}}),$$

where the subscript $\cdot_{\mathcal{H}}$ denotes the horizontal part of the vector. This metric makes π into a Riemannian submersion. Explicitly,

$$\bar{\mathcal{G}}^{\mathrm{fin}}_{(V,m)}((X, \xi m), (X, \xi m)) = m(\mathrm{tr}(VSS) + \xi^2),$$

where S is a symmetric $n \times n$ matrix that is a solution to the continuous Lyapunov equation given by $X = SV + VS$. The finite-dimensional metric so constructed is simply the cone metric built from the "balanced" case described in [14].

Let us now compute the Legendre transform. The dual variable P to $\dot{V} = X$ is given by

$$\langle P, \delta\dot{V} \rangle = \frac{d}{d\epsilon} \frac{1}{2} \bar{\mathcal{G}}^{\mathrm{fin}}_{(V,m)}((\dot{V}_\epsilon, \dot{m}), (\dot{V}_\epsilon, \dot{m})) = \frac{m}{2} \, \mathrm{tr}(\Sigma(\delta S \, S + S \, \delta S))$$

$$= \frac{m}{2} \, \mathrm{tr}((V\delta S + \delta S V)S) = \frac{m}{2} \, \mathrm{tr}(S\delta\dot{V}),$$

where for δS we have

$$\delta\dot{V} = \delta S V + V \delta S.$$

Thus, the dual variable is $P = mS/2$. The dual variable for m is ξ. This gives the Hamiltonian

$$H(V, m, P, \xi) = \frac{\mathrm{tr}(VPP)}{2m} + \frac{1}{2}m\xi^2.$$

This gives the Hamiltonian form of the geodesic equations on $T^*(\mathbf{Sym}_+(n) \times \mathbb{R}_+)$ as

$$\dot{V} = \frac{2}{m}(PV + VP), \qquad\qquad \dot{m} = \xi m$$

$$\dot{P} = -\frac{P^2}{2m}, \qquad\qquad \dot{\xi} = \frac{1}{2}\left(\frac{\operatorname{tr}(\upsilon P^2)}{m^2} - \xi^2\right)$$

Remark 2.17. Note that here the need for all four equations, as opposed to Theorem 2.7 where we only have two equations, arises from the observation in Remark 2.15 that we need two parameters to describe the unscaled Gaussian distributions.

2.5 Affine transformations and Gaussians with nonzero means

It turns out that considering the group of affine transformations $\mathrm{GL}(n) \ltimes \mathbb{R}^n \subset \mathrm{Diff}(\mathbb{R}^n)$ acting on Gaussians with arbitrary (not necessarily zero) means does not essentially change the above picture. While the group extension is semi-direct, its metric extension is a direct product, provided that a reference Gaussian is $\eta = p(x, \Sigma)\,dx$ with mean $\mu = 0$.

Remark 2.18. For a more general reference Gaussian measure the metric accumulates the following terms:

$$\mathcal{G}^{\mathrm{aff}}_{(A,b,m)}((\dot{A}, \dot{b}, \dot{m}), (\dot{A}, \dot{b}, \dot{m})) = m\left[\operatorname{tr}(\dot{A}\Sigma\dot{A}^\top) + \|\dot{A}\mu\|^2 + 2\langle\dot{A}\dot{b}, \mu\rangle + \|\dot{b}\|^2\right] + \frac{\dot{m}^2}{m},$$

and it descends to the metric $\bar{\mathcal{G}}^{\mathrm{aff}}$ on $(\mathrm{GL}(n) \ltimes \mathbb{R}^n) \times \mathbb{R}_+$ given by:

$$\bar{\mathcal{G}}^{\mathrm{aff}}_{(U,\upsilon,m)}((X, y, a), (W, z, b)) =$$
$$m\left[\operatorname{tr}(XU^{1/2}\Sigma U^{1/2}W^\top) + \langle XU^{1/2}\Sigma^{1/2}\mu, WU^{1/2}\Sigma^{1/2}\mu\rangle\right.$$
$$\left. + \langle XU^{1/2}\Sigma^{1/2}z, \mu\rangle + \langle WU^{1/2}\Sigma^{1/2}y, \mu\rangle + \langle y, z\rangle\right] + \frac{ab}{m}.$$

Note that if $\mu = 0$, then several terms vanish, and one is left with the product metric of $(\mathbf{Sym}_+(n) \times \mathbb{R}^n) \times \mathbb{R}_+$.

This implies that the geodesics between two Gaussian densities with different means are the pushforwards of measures by affine transformations, which decompose into the uniform motion between the centers of the two Gaussian densities and the $\mathrm{GL}(n)$ transformation with the fixed center.

Remark 2.19. The explicit geodesics for $\mathbf{Sym}_+(n)$ with the Wasserstein metric are given in McCann [13]. In particular, for $U, V \in \mathbf{Sym}_+(n)$, define

$$T = U^{1/2}(U^{1/2}VU^{1/2})^{-1/2}U^{1/2} \in \mathbf{Sym}_+(n),$$

and then $W(t) = [(1 - t)E + tT]V[(1 - t)E + tT]$ is a geodesic between U and V. In our case, if the reference measure is of mean zero ($\mu = 0$) the geodesics in the *balanced* affine extension are those of the product $\mathbf{Sym}_+(n) \times \mathbb{R}^n$. The geodesics in the unbalanced case are those of the conical extension $\mathbf{Sym}_+(n) \times \mathbb{R}^n \times \mathbb{R}_+$.

The sectional curvatures of $\mathbf{Sym}_+(n)$ with the Wasserstein metric are well understood (see [20]) and are known to be non-negative. Hence, in the case $\mu = 0$ the affine and conical extensions also have non-negative sectional curvatures.

3 A "Large" Extension for UOT
3.1 The "large" group, metric, and the geodesic equations

A more "classical" approach to unbalanced optimal transport involves the following large semi-direct extension of the group $\mathrm{Diff}(M)$ of all diffeomorphisms of a manifold by means of the space of smooth

functions; see, for example, [22,§3.2.2]. Namely, the semi-direct product $G = \mathrm{Diff}(M) \ltimes C^\infty_+(M)$ acts on $\mathrm{Vol}(M)$ by $(\varphi, \lambda) \cdot \varrho = \varphi_*(\lambda\varrho)$, that is, diffeomorphisms act on densities by changes of coordinates, while functions adjust the obtained density point-wise. Let $\mu \in \mathrm{Vol}(M)$ denote the reference volume form. Then we get a projection $\Pi \colon G \to \mathrm{Vol}(M)$ by the action on μ.

Lemma 3.1. The vertical bundle is given by

$$\mathcal{V}_{(\varphi,\lambda)} = \{(v \circ \varphi, \frac{\varphi^*(L_v\varrho)}{\mu}) \mid v \in \mathfrak{X}(M)\} \simeq \mathfrak{X}(M),$$

where $\varrho = \varphi_*(\lambda\mu)$.

Proof. A curve $(\varphi(t), \lambda(t))$ belongs to the fiber of $\varrho \in \mathrm{Vol}(M)$ iff $\varphi(t)_*(\lambda(t)\mu) = \varrho$ for all t. Equivalently,

$$\lambda(t) = \frac{\varphi(t)^*\varrho}{\mu}.$$

By differentiating this relation we get the result. ∎

Remark 3.2. This description of $\mathcal{V}$ is equivalent to the one given by Vialard [22] as

$$\ker d\pi(\varphi, \sqrt{\mathrm{Jac}\,\varphi}) = \left\{ \left(v, \frac{\mathrm{div}\,v}{2}\right) \circ (\varphi, \sqrt{\mathrm{Jac}\,\varphi}) \,\middle|\, v \in \mathfrak{X}(M) \right\}$$

The relation is $\varrho = \mathrm{Jac}(\varphi)\mu$, where the square root appears if one passes from volume forms to *half-densities*, that is, geometric objects that transform as the square root of a volume form.

Consider now the Riemannian metric on G studied in [5, 11, 18, 22] and given by

$$\mathcal{G}^{\mathrm{big}}_{(\varphi,\lambda)}((\dot\varphi, \dot\lambda), (\dot\varphi, \dot\lambda)) = \int_M |\dot\varphi|^2 \lambda\mu + \int_M \frac{\dot\lambda^2}{\lambda}\mu. \tag{10}$$

Lemma 3.3. The horizontal bundle of the metric $\mathcal{G}^{\mathrm{big}}$ (10) is

$$\mathcal{H}_{(\varphi,\lambda)} = \{(\nabla\theta \circ \varphi, \lambda(\theta \circ \varphi)) \mid \theta \in C^\infty(M)\} \simeq C^\infty(M).$$

Proof. Any element in $T_{(\varphi,\lambda)}G$ can be written $(u \circ \varphi, \lambda(\theta \circ \varphi))$. Suppose that $(u \circ \varphi, \lambda(\theta \circ \varphi)) \in \mathcal{H}_{(\varphi,\lambda)}$. Since for all $v \in \mathfrak{X}(M)$ the pairs $(v \circ \varphi, \varphi^*(L_v\varrho)/\mu)$ span the vertical space $\mathcal{V}_{(\varphi,\lambda)}$, we have that

$$0 = \mathcal{G}^{\mathrm{big}}_{(\varphi,\lambda)}\left(\left(v \circ \varphi, \frac{\varphi^*(L_v\varrho)}{\mu}\right), (u \circ \varphi, \lambda(\theta \circ \varphi))\right)$$

$$= \int_M \langle v \circ \varphi, u \circ \varphi \rangle \lambda\mu + \int_M (\theta \circ \varphi)\varphi^*(L_v\varrho)$$

$$= \int_M \langle v, u \rangle \varphi_*(\lambda\mu) + \int_M \theta L_v\varrho$$

$$= \int_M \langle v, u \rangle \varrho - \int_M \langle v, \nabla\theta \rangle \varrho = \int_M \langle v, u - \nabla\theta \rangle \varrho.$$

The latter integral vanishes for any $v \in \mathfrak{X}(M)$ if and only if $u = \nabla\theta$, which concludes the proof. ∎

Theorem 3.4 (cf. [22]). The metric $\mathcal{G}^{\mathrm{big}}$ given by (10) projects as a Riemannian submersion to the metric $\bar{\mathcal{G}}^{\mathrm{big}}$ on $\mathrm{Vol}(M)$ given by

$$\bar{\mathcal{G}}^{\mathrm{big}}_\varrho(\dot\varrho, \dot\varrho) = \frac{1}{2} \int_M \left(|\nabla\theta|^2 + \theta^2\right)\varrho, \qquad \dot\rho = -\mathrm{div}(\rho\nabla\theta) + \rho\theta.$$

The variable $\theta \in C^\infty(M)$ is Legendre-dual to $\dot{\varrho}$, that is, the Hamiltonian corresponding to the metric is

$$H(\varrho, \theta) = \frac{1}{2} \int_M \left(|\nabla \theta|^2 + \theta^2 \right) \varrho. \tag{11}$$

The equations of geodesics (in Hamiltonian form) are

$$\dot{\rho} = -\operatorname{div}(\rho \nabla \theta) + \rho \theta$$

$$\dot{\theta} = -\frac{1}{2} |\nabla \theta|^2 - \frac{\theta^2}{2}.$$

Proof. The proof of this follows similarly to Theorem 2.6. Hamilton's equations are $\dot{\varrho} = \delta H / \delta \theta$ and $\dot{\theta} = -\delta H / \delta \varrho$. Given a variation $\theta_\epsilon = \theta + \epsilon \, \delta \theta$, note:

$$\frac{d}{d\epsilon} \bigg|_{\epsilon=0} H(\varrho, \theta_\epsilon) = \int_M \rho (\nabla \theta \cdot \nabla(\delta \theta) + \theta \, \delta \theta) \, \mu$$

$$= \int_M \left(-\operatorname{div}(\rho \nabla \theta) + \rho \theta \right) \delta \theta \, \mu,$$

and so $\dot{\rho} = -\operatorname{div}(\rho \nabla \theta) + \rho \theta$.

Similarly, considering a variation $\varrho_\epsilon = \varrho + \epsilon \, \delta \varrho$, we see

$$\frac{d}{d\epsilon} \bigg|_{\epsilon=0} H(\varrho_\epsilon, \theta) = \int_M \frac{1}{2} \left(|\nabla \theta|^2 + \theta^2 \right) \delta \varrho,$$

and so we immediately get $\dot{\theta} = -|\nabla \theta|^2 / 2 - \theta^2 / 2$. ∎

Remark 3.5. The metric $\bar{\mathcal{G}}^{\mathrm{big}}$ in Theorem 3.4 can be interpreted as an interpolation between Wasserstein–Otto and Fisher–Rao, but not a convex combination of the Riemannian metric tensors (see Remark 4.2 below). One way of understanding the relation is the following: the Wasserstein–Otto part of the metric depends on the finite-dimensional metric g on M, but the second term does not. Thus, let us introduce a parameter β by making the replacement $g \mapsto \beta g$. Then, as $\beta \to \infty$, we recover the Fisher–Rao metric (indeed, in the Hamiltonian (11) the term with $\nabla \theta \to 0$ for the metric βg and only the second term remains, which corresponds to the Fisher–Rao metric). On the other hand, as $\beta \to 0$ we recover the (scaled) Wasserstein–Otto metric (represented by the first term in the Hamiltonian). Thus, this mixed metric behaves as Fisher–Rao on small scales, but as Wasserstein–Otto on large scales.

Remark 3.6. The metric in Theorem 3.4 lifted to a metric on $\operatorname{Diff}(M) \times \mathbb{R}_+$ is given by

$$\langle (v \circ \varphi, \dot{\lambda}), (v \circ \varphi, \dot{\lambda}) \rangle_{(\varphi, \lambda)} = \int_M \left(|\nabla \theta|^2 + \theta^2 \right) \varrho,$$

where $\varrho = \lambda \varphi_* \mu$ and $\theta \in C^\infty(M)$ is the solution to the equation

$$-\operatorname{div}(\rho \nabla \theta) + \rho \theta = -\operatorname{div}(\rho v) + \frac{\rho \dot{\lambda}}{\lambda}.$$

Notice, however, that θ is somewhat difficult to find, as it requires the solution of a non-local equation. The "small" extension discussed above does not encounter this difficulty. This shows that *the metric of the simple UOT is not a restriction* of the noticeably more complicated metric in Theorem 3.4.

3.2 Interrelation between the small and large extensions

We discussed above, in Remark 2.10, that in the simple UOT one can introduce a localized "production function" $f \geq 0$, which leads to the evolution of density ρ given by

$$\dot{\rho} = -\operatorname{div}(\rho \nabla \theta) + \xi f \rho.$$

Furthermore, one can consider a model with several production functions $f_1, \ldots, f_k$, each with an independent coefficient, which are optimised together. This way one can view this adjusted UOT as an approximation of the unbalanced transport corresponding to the large extension and the metric $\bar{\mathcal{G}}^{\mathrm{big}}$ on Vol(M), cf. Theorem 3.4 and see [22].

   Indeed, for several production functions the Hamiltonian (8) becomes

$$H(\varrho, \theta) = \frac{1}{2} \int_M |\nabla \theta|^2 \varrho + \sum_{i=1}^k \frac{1}{2m} \left(\int_M \theta f_i \varrho \right)^2,$$

where m is the total mass and $\xi_i m = \int_M \theta f_i \varrho$. Then the evolution of ρ is given by

$$\dot{\rho} = -\operatorname{div}(\rho \nabla \theta) + \rho \sum_{i=1}^k \xi_i f_i$$

where the last term $\rho \sum_1^k \xi_i f_i$ can be regarded as a finite-dimensional replacement ("approximation") of the term $\rho \theta$ with a function $\theta \in C^\infty(M)$ in the evolution of density

$$\dot{\rho} = -\operatorname{div}(\rho \nabla \theta) + \rho \theta$$

given in Theorem 3.4. Thus, as the number k of fixed production functions goes to infinity, one recovers the problem of optimal transport with variable production of density over M. Such an approximation can be useful for numerical modeling.

3.3 A finite-dimensional version of the large extension

Consider the finite-dimensional group $\mathbf{GL}(n) \ltimes \mathbf{Sym}(n) \subset \operatorname{Diff}(\mathbb{R}^n) \ltimes C_+^\infty(\mathbb{R}^n)$, where $\mathbf{Sym}(n)$ is the additive space of symmetric $n \times n$ matrices (or equivalently, the corresponding quadratic forms on $\mathbb{R}^n$), on which linear transformations act by the variable change.

   We regard $\mathbf{Sym}(n)$ as a subset of $C_+^\infty(\mathbb{R}^n)$ by using the map

$$E = \{x \mapsto \exp(x^\top S x) \mid S \in \mathbf{Sym}(n)\} \subset C_+^\infty(\mathbb{R}^n).$$

This way the addition group of symmetric matrices $\mathbf{Sym}(n)$ becomes a multiplication subgroup of positive functions $C_+^\infty(\mathbb{R}^n)$. An advantage of this approach is that the Riemannian submersion can be restricted to the finite-dimensional model, where the group $\mathbf{GL}(n) \ltimes E$ acts on $E \subset \operatorname{Vol}(M)$. Here an element $(A, S) \in \mathbf{GL}(n) \ltimes E$ acts naturally on a density $p(x, \Sigma)$ by changing variables and bringing the quadratic into the exponential:

$$(A, S) : p(x, \Sigma) \mapsto \widetilde{p}(x, \Sigma) := \exp(x^\top S x)\, p(x, A^\top \Sigma A)$$

$$= \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} x^\top ((A^\top \Sigma A)^{-1} - 2S) x \right).$$

The drawback is that even if $p(x, \Sigma)$ is a Gaussian density and S is also positive-definite, the symmetric matrix $(A^\top \Sigma A)^{-1} - 2S$ might not be positive-definite! This means that the total volume of the density $\widetilde{p}(x, \Sigma) = \exp(x^\top S x)\, p(x, A^\top \Sigma A)$ in $\mathbb{R}^n$ might be infinite. Thus, one has to consider a restricted orbit of this action, constrained by the condition of positivity of the matrix $(A^\top \Sigma A)^{-1} - 2S$.

Note that the required positivity is automatically satisfied and does not constrain anything in the infinite-dimensional setting of the space of L^2 densities Vol(M). Also this constraint is not required in the finite-dimensional simple 1D conical extension described in Section 2.4.

4 Other Versions of the Unbalanced Transport Metric
4.1 A "small" extension with a divergence term

Consider now the Riemannian metric on Diff(M) $\times \mathbb{R}_+$ given by

$$\mathcal{G}^{\text{div}}_{(\varphi,\lambda)}\left((\upsilon \circ \varphi, \xi\lambda),(\upsilon \circ \varphi, \xi\lambda)\right) = \frac{1}{2}\int_M \left(|\upsilon|^2 + \frac{\text{div}(\rho\upsilon)^2}{\rho^2}\right)\rho + \lambda\xi^2. \tag{12}$$

Notice the similarity of $\mathcal{G}^{\text{div}}$ with $\mathcal{G}$ given by (2): it is the same metric supplemented by the divergence term. In particular, on vertical vectors it is exactly the same metric. Thus, the horizontal bundle is the same as in Lemma 2.3.

Consider now the metric $\bar{\mathcal{G}}^{\text{div}}$ on Vol(M) given by

$$\bar{\mathcal{G}}^{\text{div}}_\varrho(\dot\varrho, \dot\varrho) = \int_M \left(|\nabla S|\varrho + \left(\frac{\dot\varrho}{\varrho}\right)^2 \varrho\right), \quad -\text{div}(\rho\nabla S) = \dot\rho - \kappa\rho .$$

Here, we think of κ as a Lagrange multiplier to ensure that the average of the right-hand side vanishes.

> **Theorem 4.1.** The projection $\pi \colon \text{Diff}(M) \times \mathbb{R}_+ \to \text{Vol}(M)$ given by $\pi(\varphi, \lambda) = \lambda\varphi_*\mu$ is a Riemannian submersion with respect to $\mathcal{G}^{\text{div}}$ and $\bar{\mathcal{G}}^{\text{div}}$.

Proof. The tangent derivative of the projection is

$$T_{(\varphi,\lambda)}(\upsilon \circ \varphi, \xi\lambda) = \xi\varrho - L_\upsilon\varrho.$$

In particular, for a horizontal vector $(\nabla\theta \circ \varphi, \int_M \theta\varrho)$ we have

$$T_{(\varphi,\lambda)}\left(\nabla\theta \circ \varphi, \int_M \theta\varrho\right) = \frac{\varrho}{\lambda}\int_M \theta\varrho - \text{div}(\rho\nabla\theta)\mu.$$

Taking this expression as $\dot\varrho$ we see from the definition of $\mathcal{G}^{\text{div}}$ that

$$-\text{div}(\rho\nabla S) = -\text{div}(\rho\nabla\theta).$$

Thus, $\nabla\theta = \nabla S$. We now plug this into the metric $\bar{\mathcal{G}}^{\text{div}}$:

$$\bar{\mathcal{G}}^{\text{div}}_\varrho\left(\frac{\varrho}{\lambda}\int_M \theta\varrho - \text{div}(\rho\nabla\theta)\mu, \frac{\varrho}{\lambda}\int_M \theta\varrho - \text{div}(\rho\nabla\theta)\mu\right) =$$

$$\int_M |\nabla\theta|\varrho + \frac{1}{\lambda}\int_M \theta\varrho \int_M \theta\varrho + \int_M \frac{\text{div}(\rho\nabla\theta)^2}{\rho}\mu =$$

$$\mathcal{G}^{\text{div}}_{(\varphi,\lambda)}\left(\left(\nabla\theta \circ \varphi, \int_M \theta\varrho\right), \left(\nabla\theta \circ \varphi, \int_M \theta\varrho\right)\right)$$

This proves the assertion. ∎

> **Remark 4.2.** The small conical extension Diff(M) $\times \mathbb{R}_+$ with metric $\mathcal{G}$ (see (2)) can be viewed as the "common ground" for the Lagrangian and Hamiltonian extensions in constructions of an unbalanced optimal transport.
> Indeed, the Hamiltonian $H(\varrho, \theta) = \frac{1}{2}\int_M \left(|\nabla\theta|^2 + \theta^2\right)\varrho$ (see (11)) expressing the metric $\bar{\mathcal{G}}^{\text{big}}$ on Vol(M) in the dual variables is the sum of two terms. The first one corresponds to the Wasserstein

metric, while the second term $\int_M \theta^2 \varrho = \int_M (\nu/\varrho)^2 \varrho$ for the density $\nu := \theta\varrho$ represents the Fisher-Rao metric. Hence, the name of the WFR metric for the semi-direct generalization of UOT developed in [5, 6, 22].

On the other hand, the metric $\mathcal{G}^{\mathrm{div}}$ on Vol(M) with an extra divergence term (see (12)) also has a WFR form, although not in the Hamiltonian, but in the Lagrangian setting: the first term is the Wasserstein metric, while the second, divergence term is the degenerate $\dot{H}^1$ contribution giving the Fisher-Rao metric on Vol(M).

4.2 Conical Fisher–Rao metrics

Consider the group Diff(M) equipped with a $\dot{H}^1$-type metric so that its projection to the space **Dens**(M) of normalized densities is equipped with the Fisher-Rao metric. It also admits the conical extension with the projection Diff(M) $\times$ $\mathbb{R}_+$ on Vol(M).

Note that since the Fisher-Rao metric on **Dens**(M) is spherical, its conical extension to Vol(M) $\supset$ **Dens**(M) is an (infinite-dimensional) positive quadrant of the (pre-Hilbert) space of highest-degree forms naturally equipped with the *flat L^2-metric*. The positive quadrant is formed by all volume forms on the manifold. The projection Diff(M) $\to$ **Dens**(M) from diffeomorphisms to volume forms is known to be a Riemannian submersion [10], and it remains a Riemannian submersion for its conical extension Diff(M) $\times$ $\mathbb{R}_+$ $\to$ Vol(M).

> **Proposition 4.3.** The space Diff(M) $\times$ $\mathbb{R}_+$ equipped with a conical $\dot{H}^1$-type metric has non-positive sectional curvatures.

Proof. Indeed, under a Riemannian submersion the sectional curvature cannot decrease [16]. Since the base manifold of all volume forms is flat (i.e., its sectional curvatures all vanish) under the projection, the sectional curvatures of the space Diff(M) $\times$ $\mathbb{R}_+$ must be negative or equal to zero. $\blacksquare$

This conical extension also admits a finite-dimensional version, extending the one in [14]. Indeed, the finite-dimensional submanifold **GL**(n) $\times$ $\mathbb{R}_+$ $\subset$ Diff($\mathbb{R}^n$) $\times$ $\mathbb{R}_+$ is totally geodesic and according to Lemma 2.13 the corresponding projection to Vol(M) is totally geodesic as well. One can expect similar matrix decompositions coming from this Riemannian submersion, extending those in [14].

A General Form of the Geodesic Equations for a Conical Extension

> **Theorem A .1.** The geodesic equations for the cone Q $\times$ $\mathbb{R}_+$ formed from a Riemannian manifold (Q, g) with the metric $r^2 g + dr^2$ are:
>
> $$\begin{cases} \nabla_{\dot{q}}\dot{q} + \frac{2}{\alpha}\dot{\alpha}\dot{q} = 0, \\ \ddot{\alpha} - g(\dot{q}, \dot{q})\alpha = 0, \end{cases}$$
>
> for a geodesic $\gamma = (q, \alpha) \in Q \times \mathbb{R}_+$.

Proof. For the cone Q $\times$ $\mathbb{R}_+$ consider the two projections:

$$Q \times \mathbb{R}_+ \xrightarrow{\ \pi\ } \mathbb{R}_+$$
$$\sigma \downarrow$$
$$Q$$

This cone can be viewed as a warped product $\mathbb{R}_+ \times_f Q$ for $f \colon \mathbb{R}_+ \to \mathbb{R}$ defined by $r \mapsto r$ and the metric defined for $v \in T_{(r,q)}\mathbb{R}_+ \times Q$ by:

$$\langle v, v \rangle_{(r,q)} = d\pi(v)^2 + f(r)^2 \, g_q(d\sigma(v), d\sigma(v)).$$

The geodesic equations for such a warped product is given for a geodesic $\gamma = (\alpha, q) \in \mathbb{R}_+ \times_f Q$ by

$$\begin{cases} \nabla_{\dot{\alpha}}\dot{\alpha} - g(\dot{q}, \dot{q})(f \circ \alpha)\nabla f = 0, \\ \nabla_{\dot{q}}\dot{q} + \frac{2}{f \circ \alpha}\frac{d(f \circ \alpha)}{dt}\dot{q} = 0, \end{cases}$$

see [16]. With our setup, $\nabla f = 1$ (with the standard metric on $\mathbb{R}_+$), and $f \circ r = r$, and the result follows. ∎

Remark A.2. In the present paper we apply the corresponding conical one-dimensional extension $r^2 g(v, v) + dr^2$ to the group of diffeomorphisms and the space of normalized densities, where $g(v, v)$ is, respectively, the L^2-metric on $\mathbf{Diff}(M)$ and the Wasserstein metric on $\mathbf{Dens}(M)$.

For an arbitrary $p \in \mathbb{R}$, the geodesic equations for a geodesic $\gamma = (q, \alpha)$ on the cone $Q \times \mathbb{R}_+$ with the metric $r^{2p} g(v, v) + dr^2$ assume the form:

$$\begin{cases} \nabla_{\dot{q}} \dot{q} + \frac{2p}{\alpha} \dot{\alpha} \dot{q} = 0, \\ \ddot{\alpha} - p r^{p-1} g(\dot{q}, \dot{q}) \alpha^p = 0, \end{cases}$$

of which Theorem A.1 is the special case of $p = 1$, while $p = 0$ corresponds to the direct product metric on the cylinder $Q \times \mathbb{R}$. One can also consider the one-parameter extensions $r^{2p} g + dr^2$ in infinite dimensions as well. Other hyperbolic and parabolic-type metrics for negative and positive values of p might be useful in problems of optimal transport whenever it is convenient to tune the mass balance.

Acknowledgments

Funding

References

1. Bauer, M., E. Hartman, and E. Klassen. "The square root normal field distance and unbalanced optimal transport." *Appl. Math. Optim.* **85**, no. 3 (2022): 35. https://doi.org/10.1007/s00245-022-09867-y.
2. Benamou, J.-D. "Numerical resolution of an "unbalanced" mass transport problem." *ESAIM: Math. Model. Numer. Anal.* **37**, no. 5 (2003): 851–68. https://doi.org/10.1051/m2an:2003058.
3. Benamou, J.-D. and Y. Brenier. "A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem." *Numer. Math.* **84**, no. 3 (2000): 375–93. https://doi.org/10.1007/s002110050002.
4. Chizat, L., G. Peyré, B. Schmitzer, and F.-X. Vialard. "Scaling algorithms for unbalanced optimal transport problems." *Math. Comp.* **87**, no. 314 (2018): 2563–609. https://doi.org/10.1090/mcom/3303.
5. Chizat, L., G. Peyré, B. Schmitzer, and F.-X. Vialard. "Unbalanced optimal transport: dynamic and Kantorovich formulations." *J. Func. Anal.* **274**, no. 11 (2018): 3090–123. https://doi.org/10.1016/j.jfa.2018.03.008.
6. Gallouët, T., R. Ghezzi, and F.-X. Vialard. Regularity theory and geometry of unbalanced optimal transport. (2021): arXiv preprint, arXiv:2112.11056.
7. Gangbo, W., W. Li, S. Osher, and M. Puthawala. "Unnormalized optimal transport." *J. Comput. Phys.* **399** (2019): 108940. https://doi.org/10.1016/j.jcp.2019.108940.
8. Hamilton, R. S. "The inverse function theorem of Nash and Moser." *Bull. Amer. Math. Soc. (N.S.)* **7**, no. 1 (1982): 65–222.
9. Khesin, B., G. Misiołek, and K. Modin. "Geometric hydrodynamics and infinite-dimensional Newton's equations." *Bull. Amer. Math. Soc.* **58**, no. 3 (2020): 377–442. https://doi.org/10.1090/bull/1728.
10. Khesin, B., J. Lenells, G. Misiołek, and S. C. Preston. "Geometry of diffeomorphism groups, complete integrability and geometric statistics." *Geom. Functional Anal.* **23**, no. 1 (2013): 334–66. https://doi.org/10.1007/s00039-013-0210-2.
11. Kondratyev, S., L. Monsaingeon, and D. Vorotnikov. "A new optimal transport distance on the space of finite Radon measures." *Adv. Differential Equations* **21**, no. 11-12 (2016): 1117–64. https://doi.org/10.57262/ade/1476369298.

12. Liero, M., A. Mielke, and G. Savaré. "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures." *Invent. Math.* **211**, no. 3 (2018): 969–1117. https://doi.org/10.1007/s00222-017-0759-8.

13. McCann, R. J. "A convexity principle for interacting gases." *Adv. Math.* **128**, no. 1 (1997): 153–79. https://doi.org/10.1006/aima.1997.1634.

14. Modin, K. "Geometry of matrix decompositions seen through optimal transport and information geometry." *J. Geom. Mech.* **9**, no. 3 (2017): 335–90. https://doi.org/10.3934/jgm.2017014.

15. Omori, H. *Infinite-dimensional Lie groups*, volume 158. American Mathematical Society, 2017.

16. O'Neill, B. *Semi-Riemannian Geometry with Applications to Relativity*. Academic Press, 1983.

17. Otto, F. "The geometry of dissipative evolution equations: the porous medium equation." *Comm. Partial Diff. Eq.* **26**, no. 1-2 (2001): 101–74. https://doi.org/10.1081/PDE-100002243.

18. Piccoli, B. and F. Rossi. "On properties of the generalized Wasserstein distance." *Arch. Rational Mech. Anal.* **222** (2016): 1339–65. https://doi.org/10.1007/s00205-016-1026-7.

19. Séjourné, T., F.-X. Vialard, and G. Peyré. *Faster Unbalanced Optimal Transport: Translation Invariant Sinkhorn and 1-d Frank-Wolfe*, 4995–5021. International Conference on Artificial Intelligence and Statistics, 2022.

20. Takatsu, A. "Wasserstein geometry of Gaussian measures." *Osaka J. Math.* **48**, no. 4 (2011): 1005–26.

21. Trouvé, A. "Diffeomorphisms groups and pattern matching in image analysis." *Int. J. Comput. Vision* **28**, no. 3 (1998): 213–21. https://doi.org/10.1023/A:1008001603737.

22. Vialard, F.-X. "Diffeomorphisms, optimal transport and applications to imaging." *Thèse d'habilitation*. Université Paris-Dauphine-PSL, 2017.

23. Villani, C. *Topics in Optimal Transportation*. AMS, 2003.